98829

Access DB# 83

SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: RITA MITRA Examiner #: 77995 Date: 7/2/02
Art Unit: 1653 Phone Number 30 605-1211 Serial Number: 09/600932
Mail Box and Bldg/Room Location: 9B01/CM1 Rm. 9B03 Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

****

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: Novel Collectin

Inventors (please provide full names): Nobutaka Wakamiya

Earliest Priority Filing Date: 1/23/1998

***For Sequence Searches Only* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.**

I would request an expedited literature search (Patent and Non-patent) for above application because it is date case. DON'T DO SEQ SEARCH The search should compass polynucleotides encoding collectin protein having calcium-dependent carbohydrate recognition domain (CRD) and collagen-like region.

Keyword:
gene, antiviral activity, antibacterial activity,
Note: Claims elected: 1, 2, 5, 6, 8, 9.

Ramesh

****

STAFF USE ONLY

Searcher: [illegible]
Searcher Phone #: [illegible]
Searcher Location:
Date Searcher Picked Up:
Date Completed: 7/3/02
Searcher Prep & Review Time:
Clerical Prep Time:
Online Time:

Type of Search

NA Sequence (#)
AA Sequence (#)
Structure (#)
Bibliographic
Litigation
Fulltext
Patent Family
Other

Vendors and cost where applicable

STN
Dialog
Questel/Orbit
Dr.Link
Lexis/Nexis
Sequence Systems
WWW/Internet
Other (specify)

PTO-1590 (1-2000)

```
=> fil hcaplus
FILE 'HCAPLUS' ENTERED AT 16:58:49 ON 08 JUL 2002
USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.
PLEASE SEE "HELP USAGETERMS" FOR DETAILS.
COPYRIGHT (C) 2002 AMERICAN CHEMICAL SOCIETY (ACS)
```



```
FILE COVERS 1907 - 8 Jul 2002  VOL 137 ISS 2
FILE LAST UPDATED: 7 Jul 2002  (20020707/ED)

 This file contains CAS Registry Numbers for easy and accurate
 substance identification.

CAS roles have been modified effective December 16, 2001. Please
check your SDI profiles to see if they need to be revised. For
information on CAS roles, enter HELP ROLES at an arrow prompt or use
the CAS Roles thesaurus (/RL field) in this file.

=>
=>

=> d stat que
L1            77 SEA FILE=REGISTRY ABB=ON  PLU=ON  COLLECTIN/BI
L2           260 SEA FILE=HCAPLUS ABB=ON  PLU=ON  L1 OR COLLECTIN
L4           223 SEA FILE=HCAPLUS ABB=ON  PLU=ON  L2 AND PROTEIN
L5            62 SEA FILE=HCAPLUS ABB=ON  PLU=ON  L2(L)(GENE OR DNA OR NUCLEIC(W
                 )ACID OR RNA)
L6            57 SEA FILE=HCAPLUS ABB=ON  PLU=ON  L5 AND L4


=>
=>

=> d ibib abs hitrn 16 1-57

L6   ANSWER 1 OF 57  HCAPLUS  COPYRIGHT 2002 ACS
ACCESSION NUMBER:        2002:506094  HCAPLUS
TITLE:                   Blebs and apoptotic bodies are B cell autoantigens
AUTHOR(S):               Cocca, Brian A.; Cline, Amy M.; Radic, Marko Z.
CORPORATE SOURCE:        Department of Molecular Sciences, University of
                         Tennessee Health Sciences Center, Memphis, TN, 38163,
                         USA
SOURCE:                  Journal of Immunology (2002), 169(1), 159-166
                         CODEN: JOIMA3; ISSN: 0022-1767
PUBLISHER:               American Association of Immunologists
DOCUMENT TYPE:           Journal
LANGUAGE:                English
AB   Mounting evidence suggests that systemic lupus erythematosus autoantigens
     are derived from apoptotic cells.  To characterize the potential
```

interactions between apoptotic cells and B cells, the D56R/S76R variant of 3H9, a murine autoantibody that binds to **DNA**, chromatin, and anionic phospholipids, was compared with DNA4/1, a human anti-**DNA** autoantibody. Flow cytometry revealed that only D56R/S76R bound to Jurkat cells treated with either of three distinct proapoptotic stimuli, Ab binding was dependent on caspase activity, and immunoreactivity developed subsequent to annexin V binding. Confocal microscopy established a structural basis for the distinct kinetics of binding. D56R/S76R preferentially bound to membrane blebs of apoptotic cells, whereas annexin V binding did not require blebs. Inhibition of ROCK I kinase, an enzyme that stimulates nuclear fragmentation and fragment distribution into blebs, significantly reduced Ab binding. Because members of the **collectin** and pentraxin families of serum **proteins** bind to blebs on apoptotic cells and assist in the clearance of cellular remains, our results suggest that Abs to blebs could affect the recognition of apoptotic cells by cells of the innate immune system and thus modify tolerance to nuclear Ags.

L6 ANSWER 2 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 2002:505750 HCAPLUS
TITLE: Complementation of pulmonary abnormalities in SP-D(-/-) mice with an SP-D/conglutinin fusion **protein**
AUTHOR(S): Zhang, Liqian; Hartshorn, Kevan L.; Crouch, Erika C.; Ikegami, Machiko; Whitsett, Jeffrey A.
CORPORATE SOURCE: Division of Pulmonary Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, 45229-3039, USA
SOURCE: Journal of Biological Chemistry (2002), 277(25), 22453-22459
CODEN: JBCHA3; ISSN: 0021-9258
PUBLISHER: American Society for Biochemistry and Molecular Biology
DOCUMENT TYPE: Journal
LANGUAGE: English
AB Surfactant **protein** D (SP-D) and serum conglutinin are closely related members of the **collectin** family of host defense lectins. Although normally synthesized at different anat. sites, both **proteins** participate in the innate immune response to microbial challenge. To discern the roles of specific domains in the function of SP-D in vivo, a fusion **protein** (SP-D/Congneck+CRD) consisting of the NH2-terminal and collagenous domains of rat SP-D (rSP-D) and the neck and carbohydrate recognition domains (CRDs) of bovine conglutinin (Cong) was expressed in the respiratory epithelium of SP-D **gene** -targeted (SP-D(-/-)) mice. While SP-D/Congneck+CRD fusion **protein** did not affect lung morphol. and surfactant phospholipid levels in the lungs of wild type mice, the chimeric **protein** substantially cor. the increased lung phospholipids in SP-D(-/-) mice. The SP-D/Congneck+CRD fusion **protein** also completely cor. defects in influenza A clearance and inhibited the exaggerated inflammatory response that occurs following viral infection. However, the chimeric **protein** did not ameliorate the ongoing lung inflammation, enhanced metalloproteinase expression, and alveolar destruction that characterize this model of SP-D deficiency. By contrast, a single arm mutant (RrSP-DSer15,20) partially restored antiviral activity but otherwise failed to rescue the deficient phenotype. Our findings directly implicate the CRDs of both SP-D and conglutinin in host defense in vivo. Our findings also strongly suggest that the mol. mechanisms underlying impaired pulmonary host defense and abnormal lipid metab. are distinct

from those that promote ongoing inflammation and the development of emphysema.

L6 ANSWER 3 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 2002:451532 HCAPLUS
TITLE: Cutting edge: the immunostimulatory activity of the lung surfactant **protein**-A involves toll-like receptor 4
AUTHOR(S): Guillot, Loic; Balloy, Viviane; McCormack, Francis X.; Golenbock, Douglas T.; Chignard, Michel; Si-Tahar, Mustapha
CORPORATE SOURCE: Unite de Defense Innee et Inflammation, Institut Pasteur, Institut National de la Sante et de la Recherche Medicale, Paris, 75015, Fr.
SOURCE: Journal of Immunology (2002), 168(12), 5989-5992
CODEN: JOIMA3; ISSN: 0022-1767
PUBLISHER: American Association of Immunologists
DOCUMENT TYPE: Journal
LANGUAGE: English
AB The **collectin** surfactant **protein**-A (SP-A) is involved in the innate host defense and the regulation of inflammatory processes in the lung. In this work we investigated the mol. mechanisms related to the immunostimulatory activity of SP-A using macrophages from C3H/HeJ mice, which carry an inactivating mutation in the Toll-like receptor (TLR)4 **gene**, and TLR4-transfected Chinese hamster ovary cells. We demonstrate that SP-A-induced activation of the NF-.kappa.B signaling pathway and up-regulation of cytokine synthesis such as TNF-.alpha. and IL-10 are critically dependent on the TLR4 functional complex. These findings support the concept that TLR4 is a pattern recognition receptor that signals in response to both foreign pathogens and endogenous host mediators.
REFERENCE COUNT: 30 THERE ARE 30 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 4 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 2002:408542 HCAPLUS
DOCUMENT NUMBER: 137:5000
TITLE: Vaccine composition comprising immunogenic determinant and **collectin** as adjuvant
INVENTOR(S): Jensenius, Jens Christian; Sjoeholm, Anders
PATENT ASSIGNEE(S): Den.
SOURCE: PCT Int. Appl., 54 pp.
CODEN: PIXXD2
DOCUMENT TYPE: Patent
LANGUAGE: English
FAMILY ACC. NUM. COUNT: 1
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 2002041913	A1	20020530	WO 2001-DK786	20011127

W: AE, AG, AL, AM, AT, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, CZ, DE, DE, DK, DK, DM, DZ, EC, EE, EE, ES, FI, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW, AM, AZ, BY, KG, KZ
RW: GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW, AT, BE, CH,

CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR, BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG
PRIORITY APPLN. INFO.: DK 2000-1785 A 20001127
AB The present invention provides vaccine compns. comprising **collectins** and immunogenic determinants. The immunogenic determinant comes from a bacterial, fungal, viral or other pathogenic antigen; and the **collectin** is SP-A, SP-D, CL43, conglutinin, CL1 and mannose binding **protein** (MBL). Furthermore, the invention describes methods of immunizing individuals with said compns. as well as the use of **collectins** for prepn. of vaccine compns.
REFERENCE COUNT: 6 THERE ARE 6 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 5 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 2002:209441 HCAPLUS
DOCUMENT NUMBER: 136:323913
TITLE: Surfactant **protein**-A-deficient mice display an exaggerated early inflammatory response to a .beta.-resistant strain of influenza A virus
AUTHOR(S): Li, Gordon; Siddiqui, Jiyauddin; Hendry, Michael; Akiyama, Jennifer; Edmondson, Jess; Brown, Cynthia; Allen, Lennell; Levitt, Stacey; Poulain, Francis; Hawgood, Samuel
CORPORATE SOURCE: Departments of Pediatrics and Cardiovascular Research Institute, University of California San Francisco, San Francisco, CA, 94118-1245, USA
SOURCE: American Journal of Respiratory Cell and Molecular Biology (2002), 26(3), 277-282
CODEN: AJRBEL; ISSN: 1044-1549
PUBLISHER: American Thoracic Society
DOCUMENT TYPE: Journal
LANGUAGE: English
AB Surfactant **protein** (SP)-A is a member of the **collectin** family of **proteins**. In vitro, SP-A binds influenza A virus (IAV), neutralizes infectivity, and enhances uptake by macrophages. SP-D also binds and neutralizes certain strains of IAV. To det. if SP-A has a role in protecting the intact animal against IAV infection, the authors inoculated **gene**-targeted SP-A-deficient mice (-/-) and littermate controls (+/+) with either saline or increasing doses of an IAV strain that binds SP-A but not SP-D. IAV was more virulent in SP-A-/- compared with +/+ mice, with a lower mean LD (LD50) and greater wt. loss during infection. SP-A-/- mice also had increased airway epithelial injury and more alveolar cellular infiltrates than +/+ mice. On Day 2, SP-A-/- mice had more neutrophils and higher MIP-2 levels in the lung than +/+ mice. Thus, the altered host response and increased susceptibility to X-79.DELTA.167 infection in SP-A-/- mice reflect a protective role for SP-A in regulating the host response to IAV. Because the recovery of virus from lung homogenates on Days 2 and 6 after inoculation was comparable in -/- and +/+ mice, the authors speculate SP-A reduces IAV virulence independently of direct viral neutralization.
REFERENCE COUNT: 34 THERE ARE 34 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 6 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 2001:798277 HCAPLUS
DOCUMENT NUMBER: 135:353796
TITLE: cDNA and **protein** sequences of novel **collectins** (CL-L2) from human and mouse and their uses for drug screening

INVENTOR(S): Wakamiya, Nobutaka; Keshi, Hiroyuki; Ohtani, Katsuki; Sakamoto, Takashi; Kishi, Yuichiro
PATENT ASSIGNEE(S): Fuso Pharmaceutical Industries, Ltd., Japan
SOURCE: PCT Int. Appl., 134 pp.
CODEN: PIXXD2
DOCUMENT TYPE: Patent
LANGUAGE: Japanese
FAMILY ACC. NUM. COUNT: 1
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 2001081401	A1	20011101	WO 2001-JP3468	20010423

W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM
RW: GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG

PRIORITY APPLN. INFO.: JP 2000-120358 A 20000421

AB The invention provides the cDNA and **protein** sequence of human and mouse **collectins** (CL-L2) and splicing derivs. of CL-L2 cloned from EST(expression sequence tags). The CL-L2s contain Gly-Xaa-Yaa repeating motif in N-terminal of the sequence and CRD domain. The purified CL-L2 provided in this invention showed carbohydrate binding activity. The invention also provides the tissue distribution of CL-L2 **genes**. The CL-L2s can be used for drug screening for identification of agonists and antagonists against CL-L2.

IT **252198-24-6 371921-21-0 371921-22-1**
371921-23-2 372025-56-4, **Collectin** CL-L2-2 (human) **372025-57-5**, 113-271-**Collectin** CL-L2-1 (human)
372025-58-6, 41-112-**Collectin** CL-L2-1 (human)
372025-61-1, **Collectin** CL-L2-2v1 (human)
372025-65-5, **Collectin** CL-L2-2v2 (human)
372025-69-9, **Collectin** CL-L2-2v3 (human)
372025-71-3, **Collectin** m-CL-L2 (Mus musculus)
372025-74-6, **Collectin** CL-L2-1v1 (human)
372025-78-0, **Collectin** CL-L2-1v3 (human)
372025-81-5, **Collectin** CL-L2-1v2 (human)
372144-36-0 372144-58-6 372144-59-7
372144-60-0, 44-112-**Collectin** CL-L2-1 (human)
372144-61-1, 68-112-**Collectin** CL-L2-1 (human)
372144-62-2, 1-40-**Collectin** CL-L2-1 (human)
RL: BOC (Biological occurrence); BSU (Biological study, unclassified); PRP (Properties); THU (Therapeutic use); BIOL (Biological study); OCCU (Occurrence); USES (Uses)
(amino acid sequence; cDNA and **protein** sequences of novel **collectins** (CL-L2) from human and mouse and their uses in diagnosis and therapeutics)

IT **372025-50-8 372025-54-2**, **DNA** (human **collectin** CL-L2-1 cDNA) **372025-55-3 372025-63-3**, **DNA** (human **collectin** CL-L2-2v1 cDNA)
372025-64-4 372025-66-6, **DNA** (human **collectin** CL-L2-2v2 cDNA) **372025-67-7**
372025-68-8, **DNA** (human **collectin** CL-L2-2v3 cDNA) **372025-70-2 372025-72-4 372025-73-5**

372025-75-7, DNA (human **collectin** CL-L2-1v1 cDNA) **372025-76-8, DNA** (human **collectin** CL-L2-1v2 cDNA) **372025-77-9 372025-79-1, DNA** (human **collectin** CL-L2-1v3 cDNA) **372025-80-4 372026-64-7 372026-65-8, DNA** (human **collectin** CL-L2-2 cDNA) **372026-66-9 372026-67-0 372026-68-1**
RL: BOC (Biological occurrence); BSU (Biological study, unclassified); PRP (Properties); THU (Therapeutic use); BIOL (Biological study); OCCU (Occurrence); USES (Uses)
(nucleotide sequence; cDNA and **protein** sequences of novel **collectins** (CL-L2) from human and mouse and their uses in diagnosis and therapeutics)

REFERENCE COUNT: 10 THERE ARE 10 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 7 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 2001:598172 HCAPLUS
DOCUMENT NUMBER: 135:176473
TITLE: Human and mouse scavenger receptor SRCL-P1
INVENTOR(S): Wakamiya, Nobutaka
PATENT ASSIGNEE(S): Fuso Pharmaceutical Industries, Ltd., Japan
SOURCE: PCT Int. Appl., 118 pp.
CODEN: PIXXD2
DOCUMENT TYPE: Patent
LANGUAGE: Japanese
FAMILY ACC. NUM. COUNT: 1
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 2001059107	A1	20010816	WO 2001-JP874	20010208

W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM
RW: GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG

PRIORITY APPLN. INFO.: JP 2000-35155 A 20000214
JP 2000-309068 A 20001010

AB Novel scavenger receptor SRCL-P1 from human and mouse having an SR structure and a **collectin**-like domain, cDNAs, recombinant expression, transgenic or knockout animal, antibodies and use in drug screening, are disclosed. Using a human placenta cDNA library, cDNA for a novel member belonging to the scavenger receptor family was cloned. Complementary **DNA** of this clone encodes a type II transmembranous glycoprotein contg. a collagen-like domain, which are typical structural characteristics of scavenger receptor class A. This **protein** also contains a C-type lectin/carbohydrate recognition domain (C-type CRD) located at the C-terminus. We designated this as Scavenger Receptor with C-type Lectin (SRCL). When SRCL-P1 were expressed in CHO cells, they were localized in the plasma membrane forming clusters on the surface. Ligand-binding studies of CHO cells expressing SRCL-P1 demonstrated their specific binding capacity in Escherichia coli and Staphylococcus aureus as well as oxidized LDL and advanced glycation end products (AGE).

REFERENCE COUNT: 8 THERE ARE 8 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 8 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 2001:567208 HCAPLUS
DOCUMENT NUMBER: 135:326846
TITLE: Prophylaxis and treatment of influenza virus infection
AUTHOR(S): Kandel, Ruth; Hartshorn, Kevan L.
CORPORATE SOURCE: Hebrew Rehabilitation Center for Aged, Harvard University School of Medicine and Section of Hematology/Oncology, Boston University School of Medicine, Boston, MA, USA
SOURCE: BioDrugs (2001), 15(5), 303-323
CODEN: BIDRF4; ISSN: 1173-8804
PUBLISHER: Adis International Ltd.
DOCUMENT TYPE: Journal; General Review
LANGUAGE: English

AB A review with refs. Influenza virus infections remain an important cause of morbidity and mortality. Furthermore, a recurrence of pandemic influenza remains a real possibility. There are now effective ways to both prevent and treat influenza. Prevention of infection is most effectively accomplished by vaccination. Vaccination with the inactivated, i.m. influenza vaccine has been clearly demonstrated to reduce serious morbidity and mortality assocd. with influenza infection, esp. in groups of patients at high risk (e.g. the elderly). However, the inactivated, i.m. vaccine does not strongly induce cell-mediated or mucosal immune responses, and protection induced by the vaccine is highly strain specific. Live, attenuated influenza vaccines administered intranasally have been studied in clin. trials and shown to elicit stronger mucosal and cell-mediated immune responses. Live, attenuated vaccines appear to be more effective for inducing protective immunity in children or the elderly than inactivated, i.m. vaccines. Addnl., novel vaccine methodologies employing conserved components of influenza virus or viral **DNA** are being developed. Preclin. studies suggest that these approaches may lead to methods of vaccination that could induce immunity against diverse strains or subtypes of influenza. Because of the limitations of vaccination, antiviral therapy continues to play an important role in the control of influenza. Two major classes of antivirals have demonstrated ability to prevent or treat influenza in clin. trials: the adamantanes and the neuraminidase inhibitors. The adamantanes (amantadine and rimantadine) have been in use for many years. They inhibit viral uncoating by blocking the proton channel activity of the influenza A viral M2 **protein**. Limitations of the adamantanes include lack of activity against influenza B, toxicity (esp. in the elderly), and the rapid development of resistance. The neuraminidase inhibitors were designed to interfere with the conserved sialic acid binding site of the viral neuraminidase and act against both influenza A and B with a high degree of specificity when administered by the oral (oseltamivir) or inhaled (zanamivir) route. The neuraminidase inhibitors have relatively low toxicity, and viral resistance to these inhibitors appears to be uncommon. Addnl. novel antivirals that target other phases of the life cycle of influenza are in preclin. development. For example, recombinant **collectins** inhibit replication of influenza by binding to the viral haemagglutinin as well as altering phagocyte responses to the virus. Recombinant techniques have been used for generation of antiviral **proteins** (e.g. modified **collectins**) or oligonucleotides. Greater understanding of the biol. of influenza viruses has already resulted in significant advances in the management of this important pathogen. Further advances in

vaccination and antiviral therapy of influenza should remain a high priority.

REFERENCE COUNT: 135 THERE ARE 135 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 9 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 2001:545709 HCAPLUS
DOCUMENT NUMBER: 135:148240
TITLE: Human nucleic acids and polypeptides
INVENTOR(S): Tang, Y. Tom; Liu, Chenghua; Asundi, Vinod; Chen, Rui-hong; Ma, Yunqing; Qian, Xiaohong B.; Ren, Feiyan; Wang, Dunrui; Wang, Jian-rui; Wang, Zhiwei; Wehrman, Tom; Xu, Chongjun; Xue, Aidong J.; Yang, Yonghong; Zhang, Jie; Zhao, Qing A.; Zhou, Ping; Goodrich, Ryle; Drmanac, Radoje T.
PATENT ASSIGNEE(S): Hyseq, Inc., USA; et al.
SOURCE: PCT Int. Appl., 10078 pp.
CODEN: PIXXD2
DOCUMENT TYPE: Patent
LANGUAGE: English
FAMILY ACC. NUM. COUNT: 65
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 2001053312	A1	20010726	WO 2000-US34263	20001226
W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM				
RW: GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG				
AU 2001025965	A5	20010731	AU 2001-25965	20001222

PRIORITY APPLN. INFO.:
US 2000-488725 A 20000121
US 2000-552317 A 20000425
US 2000-598042 A 20000709
US 2000-620312 A 20000719
US 2000-653450 A 20000803
US 2000-662191 A 20000915
US 2000-693036 A 20001019
US 2000-727344 A 20001129
US 1999-471275 A 19991223
WO 2000-US35190 W 20001222

AB The present invention provides 1768 novel nucleic acids, novel polypeptide sequences encoded by these nucleic acids, and uses thereof. A plurality of novel nucleic acids were obtained from cDNA libraries prepd. from various human tissues and in some cases from a genomic library derived from human chromosomes using std. PCR, sequencing by hybridization (SBH) sequence signature anal., and Sanger sequencing techniques. The contigs or nucleic acids of the present invention were assembled using an EST sequence as a seed, with a recursive algorithm used to extend the seed EST into an extended assemblage by pulling addnl. sequences from different databases that belong to this assemblage. Full-length gene cDNA sequences and their corresponding protein sequences were generated from the assemblage.

REFERENCE COUNT: 6 THERE ARE 6 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 10 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 2001:351881 HCAPLUS
DOCUMENT NUMBER: 135:74180
TITLE: Surfactant **proteins** and cell markers in the respiratory epithelium of the amphibian, Ambystoma mexicanum
AUTHOR(S): Miller, L.-A. D.; Wert, S. E.; Whitsett, J. A.
CORPORATE SOURCE: Division of Pulmonary Biology, Children's Hospital Medical Center, Cincinnati, OH, 45229-3039, USA
SOURCE: Comparative Biochemistry and Physiology, Part A: Molecular & Integrative Physiology (2001), 129A(1), 141-149
CODEN: CBPAB5; ISSN: 1095-6433
PUBLISHER: Elsevier Science Inc.
DOCUMENT TYPE: Journal
LANGUAGE: English
AB The respiratory tract is lined by diverse epithelial cell types whose morphol., **gene** expression and functions are highly specialized along the cephalo-caudal axis of the lung. Pulmonary gas exchange, surface tension redn., host defense, fluid and electrolyte transport are functions shared by various vertebrate species, each organism facing similar requirements for adaptation to air breathing. Consistent with this concept, the authors have identified distinct respiratory epithelial cell populations in the amphibian, Ambystoma mexicanum, using morphol., histochem. and immunochem. techniques. Thyroid transcription factor-1 (TTF-1), a homeodomain nuclear transcription factor crit. to lung formation, and surfactant **protein** B (SP-B), an amphipathic polypeptide required for surfactant function, were detected in the peripheral respiratory epithelial cells of the axolotl lung, in cells with characteristics of Type II alveolar epithelial cells in mammals. .beta.-Tubulin and carbohydrate staining identified distinct subsets of ciliated and goblet cells. SP-D, a member of the **collectin** family of innate host defense **proteins**, was also detected in peripheral epithelial cells of the axolotl lung. Pulmonary surfactant and host defense **proteins** are shared across diverse phyla supporting the concept that pulmonary structure and function have evolved from common ancestors.
REFERENCE COUNT: 29 THERE ARE 29 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 11 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 2001:348423 HCAPLUS
DOCUMENT NUMBER: 135:120347
TITLE: Clinical biological and genetic heterogeneity of the inborn errors of pulmonary surfactant metabolism
AUTHOR(S): Tredano, Mohammed; De Blic, Jacques; Griese, Matthias; Fournet, Jean-Christophe; Elion, Jacques; Bahuau, Michel
CORPORATE SOURCE: Service de Biochimie et Biologie Moleculaire, Hopital d'Enfants Armand-Trousseau, Paris, Fr.
SOURCE: Clinical Chemistry and Laboratory Medicine (2001), 39(2), 90-108
CODEN: CCLMFW; ISSN: 1434-6621
PUBLISHER: Walter de Gruyter GmbH & Co. KG
DOCUMENT TYPE: Journal; General Review
LANGUAGE: English

AB A review with 87 refs. Pulmonary surfactant is a multimol. complex located at the air-water interface within the alveolus to which a range of phys. (surface-active properties) and immune functions has been assigned. This complex consists of a surface-active lipid layer (consisting mainly of phospholipids), and of an aq. subphase. From discrete surfactant sub-fractions one can isolate strongly hydrophobic surfactant **proteins** B (SP-B) and C (SP-C) as well as **collectins** SP-A and SP-D, which were shown to have specific structural, metabolic, or immune properties. Inborn or acquired abnormalities of the surfactant, qual. or quant. in nature, account for a no. of human diseases. Beside hyaline membrane disease of the preterm neonate, a cluster of hereditary or acquired lung diseases has been characterized by periodic acid-Schiff-pos. material filling the alveoli. From this heterogeneous nosol. group, at least two discrete entities presently emerge. The first is the SP-B deficiency, in which an essentially proteinaceous material is stored within the alveoli, and which represents an autosomal recessive Mendelian entity linked to the SFTPB **gene** (MIM 1786640). The disease usually generally entails neonatal respiratory distress with rapid fatal outcome, although partial or transient deficiencies have also been obsd. The second is alveolar proteinosis, characterized by the storage of a mixed **protein** and lipid material, which constitutes a relatively heterogeneous clin. and biol. syndrome, esp. with regard to age at onset (from the neonate through to adulthood) as well as the severity of assocd. signs. Murine models, with a targeted mutation of the **gene** encoding granulocyte macrophage colony-stimulating factor (GM-CSF) (Csfgm) or the .beta. subunit of its receptor (II3rb1) support the hypothesis of an abnormality of surfactant turnover in which the alveolar macrophage is a key player. Apart from SP-B deficiency, in which a near-consensus diagnostic chart can be designed, the ascertainment of other abnormalities of surfactant metab. is not straightforward. The disentanglement of this disease cluster is however essential to propose specific therapeutic procedures: repeated broncho-alveolar lavages, GM-CSF replacement, bone marrow grafting or lung transplantation.

REFERENCE COUNT: 87 THERE ARE 87 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 12 OF 57 HCAPLUS COPYRIGHT 2002 ACS

ACCESSION NUMBER: 2001:285788 HCAPLUS

DOCUMENT NUMBER: 135:270691

TITLE: Clinical, biological and genetic heterogeneity of the inborn errors of pulmonary surfactant metabolism: SP-B deficiency and alveolar proteinosis

AUTHOR(S): Tredano, M.; De Blic, J.; Griese, M.; Fournet, J.-C.; Elion, J.; Bahuau, M.

CORPORATE SOURCE: Service de biochimie et biologie moleculaire, Hopital d'enfants Armand-Trousseau, Paris, 75571, Fr.

SOURCE: Annales de Biologie Clinique (2001), 59(2), 131-148 CODEN: ABCLAI; ISSN: 0003-3898

PUBLISHER: John Libbey Eurotext

DOCUMENT TYPE: Journal; General Review

LANGUAGE: French

AB A review, with 83 refs. Pulmonary surfactant is a multimol. complex located at the air-water interface within the alveolus and to which a bulk of functions has been assigned, phys. (surface-active properties) as well as immune or depurant. This complex consists of a surface active lipid layer (mainly phospholipids), and of an aq. subphase. From discrete surfactant sub-fractions, one can isolate very hydrophobic **proteins** SP-B and SP-C as well as the **collectins** SP-A and SP-D, which were shown to have structural, metabolic, or defensive

properties. Inborn or acquired abnormalities of surfactant, qual. or quant. in nature, account for a no. human diseases. Beside hyaline membrane disease of the preterm neonate, a cluster of hereditary or acquired lung diseases have been characterized by the storage of periodic acid Schiff-pos. material filling the alveoli. From this heterogeneous nosol. bulk, at least two discrete entities presently seem to emerge: (1) SP-B deficiency, in which an essentially proteinaceous material is stored within the alveoli, and which is a bona fide autosomal recessive Mendelian entity linked to the SFTPB **gene** (MIM 1786640), generally entailing neonatal respiratory distress with rapid fatal outcome, although partial or transient deficiencies have also been obsd.; (2) alveolar proteinosis, characterized by the storage of a mixed, **protein** and lipid material, and which constitutes a relatively heterogeneous clin. biol. syndrome, with regards to age at onset (from the neonate through to adulthood) as well as the severity of assocd. signs. Murine models with a targeted mutation of the **gene** encoding GM-CSF (Csfgm) or the beta subunit of its receptor (Il3rbl) support the hypothesis of an abnormality of surfactant turnover in which the alveolar macrophage would be a key player. Beside SP-B deficiency, in which a near-consensus diagnostic chart can be designed, the ascertainment of other abnormalities of surfactant metab. is not straightforward. The disentanglement of this disease cluster is however essential, with aim to propose differentiated therapeutic procedure: repeated bronchoalveolar lavages, GM-CSF replacement, bone marrow grafting or lung transplantation.

REFERENCE COUNT: 83 THERE ARE 83 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 13 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 2001:79829 HCAPLUS
DOCUMENT NUMBER: 135:119959
TITLE: Structures and functions of mammalian **collectins**
AUTHOR(S): Kishore, Uday; Reid, Kenneth B. M.
CORPORATE SOURCE: Institute of Molecular Medicine, University of Oxford, Headington, Oxford, OX3 9DS, UK
SOURCE: Results and Problems in Cell Differentiation (2000), 33(Mammalian Carbohydrate Recognition Systems), 225-248
CODEN: RCLDAT; ISSN: 0080-1844
PUBLISHER: Springer-Verlag
DOCUMENT TYPE: Journal; General Review
LANGUAGE: English

AB A review with 83 refs. Topics discussed include the mol. structure and assembly of mannose-binding lectin (MBL); biol. functions of MBL; interaction of MBL with microorganisms; **gene** organization and genetics of MBL; crystal structure of trimeric carbohydrate recognition domains (CRDs) of MBL; surfactant **protein** A (SP-A) superstructure and assembly; SP-A **gene** and genomic organization; SP-A carbohydrate interaction; SP-A phospholipid interactions; SP-A type II cell interaction; interaction of SP-A with phagocytes; interaction of SP-A with pathogens and allergens; mol. structure and assembly of SP-D; interaction of SP-D with carbohydrate and lipid ligands; interaction of SP-D with pathogens and allergens; SP-D **gene** organization and genetics; SP-D crystal structure; cell surface receptors for **collectins**; SP-A and SP-D **gene** knock-out mice; SP-A and SP-D in human diseases; and bovine **collectins**.

REFERENCE COUNT: 83 THERE ARE 83 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 14 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 2000:861803 HCAPLUS
DOCUMENT NUMBER: 134:26110
TITLE: Human secreted and transmembrane polypeptides and nucleic acids encoding the same
INVENTOR(S): Ashkenazi, Avi J.; Baker, Kevin P.; Botstein, David; Desnoyers, Luc; Eaton, Dan L.; Ferrara, Napoleone; Fong, Sherman; Gerber, Hanspeter; Gerritsen, Mary E.; Goddard, Audrey; Godowski, Paul J.; Grimaldi, Christopher J.; Gurney, Austin L.; Kljavin, Ivar J.; Napier, Mary A.; Pan, James; Paoni, Nicholas F.; Roy, Margaret Ann; Stewart, Timothy A.; Tumas, Daniel; Watanabe, Colin K.; Williams, P. Mickey; Wood, William I.; Zhang, Zemin
PATENT ASSIGNEE(S): Genentech, Inc., USA
SOURCE: PCT Int. Appl., 935 pp.
CODEN: PIXXD2
DOCUMENT TYPE: Patent
LANGUAGE: English
FAMILY ACC. NUM. COUNT: 71
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 2000073454	A1	20001207	WO 2000-US8439	20000330
W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM				
RW: GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG				
WO 9963088	A2	19991209	WO 1999-US12252	19990602
WO 9963088	A3	20010329		
W: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM				
RW: GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG				
WO 2000015796	A2	20000323	WO 1999-US21090	19990915
WO 2000015796	A3	20000824		
W: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM				
RW: GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG				
WO 2000015797	A2	20000323	WO 1999-US21547	19990915
WO 2000015797	A3	20000824		
W: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU,				

CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL,
IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD,
MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK,
SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY,
KG, KZ, MD, RU, TJ, TM

RW: GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE,
DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF,
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG

WO 2000032221 A2 20000608 WO 1999-US28313 19991130

WO 2000032221 A3 20001123

W: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU,
CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL,
IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD,
MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK,
SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ,
BY, KG, KZ, MD, RU, TJ, TM

RW: GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE,
DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF,
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG

WO 2000032776 A2 20000608 WO 1999-US28301 19991201

WO 2000032776 A3 20001228

W: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU,
CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL,
IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD,
MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK,
SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ,
BY, KG, KZ, MD, RU, TJ, TM

RW: GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE,
DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF,
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG

WO 2000037640 A2 20000629 WO 1999-US30095 19991216

WO 2000037640 A3 20001228

W: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU,
CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL,
IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD,
MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK,
SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY,
KG, KZ, MD, RU, TJ, TM

RW: GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE,
DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF,
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG

WO 2000075316 A1 20001214 WO 1999-US30911 19991220

W: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU,
CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL,
IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD,
MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK,
SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY,
KG, KZ, MD, RU, TJ, TM

RW: GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE,
DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF,
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG

WO 2000053753 A2 20000914 WO 2000-US219 20000105

WO 2000053753 A3 20010419

W: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU,
CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL,
IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD,
MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK,
SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ,
BY, KG, KZ, MD, RU, TJ, TM

RW: GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE,
DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF,
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG

WO 2000053755 A2 20000914 WO 2000-US376 20000106
WO 2000053755 A3 20010208
W: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU,
CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL,
IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD,
MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK,
SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ,
BY, KG, KZ, MD, RU, TJ, TM
RW: GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE,
DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF,
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG

WO 2001053486 A1 20010726 WO 2000-US3565 20000211
W: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU,
CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL,
IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA,
MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI,
SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM,
AZ, BY, KG, KZ, MD, RU, TJ, TM
RW: GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE,
DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF,
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG

WO 2000053756 A2 20000914 WO 2000-US4341 20000218
WO 2000053756 A3 20010201
W: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU,
CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL,
IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA,
MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI,
SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM,
AZ, BY, KG, KZ, MD, RU, TJ, TM
RW: GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE,
DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF,
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG

WO 2001004311 A1 20010118 WO 2000-US4414 20000222
W: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU,
CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL,
IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA,
MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI,
SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM,
AZ, BY, KG, KZ, MD, RU, TJ, TM
RW: GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE,
DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF,
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG

WO 2000075327 A1 20001214 WO 2000-US4914 20000224
W: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU,
CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL,
IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA,
MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI,
SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM,
AZ, BY, KG, KZ, MD, RU, TJ, TM
RW: GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE,
DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF,
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG

EP 1210418 A1 20020605 EP 2000-916675 20000330
R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT,
IE, SI, LT, LV, FI, RO, MK, CY, AL
EP 1212417 A2 20020612 EP 2000-939307 20000517

R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, SI, LT, LV, FI, RO, MK, CY, AL

WO 2000077037 A2 20001221 WO 2000-US14042 20000522

WO 2000077037 A3 20020228

W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM

RW: GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG

EP 1185648 A2 20020313 EP 2000-941164 20000530

R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, SI, LT, LV, FI, RO

WO 2001009327 A2 20010208 WO 2000-US20710 20000728

WO 2001009327 A3 20010802

W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM

RW: GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG

WO 2001040464 A1 20010607 WO 2000-US22031 20000811

WO 2001040464 C1 20010628

W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM

RW: GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG

WO 2001016319 A2 20010308 WO 2000-US23522 20000823

WO 2001016319 A3 20011004

W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM

RW: GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG

EP 1208201 A2 20020529 EP 2000-959474 20000823

R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, SI, LT, LV, FI, RO, MK, CY, AL

WO 2001016318 A2 20010308 WO 2000-US23328 20000824

W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM

RW: GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG

EP 1208202 A2 20020529 EP 2000-964919 20000824

R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, SI, LT, LV, FI, RO, MK, CY, AL

WO 2001049715 A2 20010712 WO 2000-US30952 20001108

WO 2001049715 A3 20020404

W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM

RW: GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG

WO 2001068848 A2 20010920 WO 2001-US6520 20010228

W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM

RW: GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG

PRIORITY APPLN. INFO.:

WO 1999-US12252 A 19990602
US 1999-141037P P 19990623
US 1999-143048P P 19990707
US 1999-144758P P 19990720
US 1999-145698P P 19990726
US 1999-146222P P 19990728
US 1999-149396P P 19990817
WO 1999-US21090 A 19990915
WO 1999-US21547 A 19990915
US 1999-158663P P 19991008
WO 1999-US28313 A 19991130
WO 1999-US28301 A 19991201
WO 1999-US30095 A 19991216
WO 1999-US30911 A 19991220
WO 2000-US219 A 20000105
WO 2000-US376 A 20000106
WO 2000-US3565 A 20000211
WO 2000-US4341 A 20000218
WO 2000-US4414 A 20000222
WO 2000-US4914 A 20000224
US 1998-87607P P 19980602
US 1998-87609P P 19980602
US 1998-87759P P 19980602
US 1998-87827P P 19980603
US 1998-88021P P 19980604
US 1998-88025P P 19980604
US 1998-88028P P 19980604
US 1998-88029P P 19980604
US 1998-88030P P 19980604
US 1998-88033P P 19980604
US 1998-88326P P 19980604
US 1998-88167P P 19980605

US 1998-88202P P 19980605
US 1998-88212P P 19980605
US 1998-88217P P 19980605
US 1998-88655P P 19980609
US 1998-88722P P 19980610
US 1998-88730P P 19980610
US 1998-88734P P 19980610
US 1998-88738P P 19980610
WO 1998-US19330 W 19980916
US 1998-100858P P 19980917
WO 1998-US19437 W 19980917
WO 1998-US25108 W 19981201
US 1998-112850P P 19981216
US 1998-113296P P 19981222
US 1999-115554P P 19990112
WO 1999-US5028 W 19990308
US 1999-123972P P 19990311
US 1999-123957P P 19990312
US 1999-126773P P 19990329
US 1999-130232P P 19990421
US 1999-131445P P 19990428
US 1999-133459P P 19990511
US 1999-134287P P 19990514
US 1999-139695P P 19990615
US 1999-140650P P 19990622
US 1999-140653P P 19990622
US 1999-145070P P 19990720
US 1999-149395P P 19990817
US 1999-151689P P 19990831
US 1999-151733P P 19990831
WO 1999-US201 W 19990901
WO 1999-US20111 W 19990901
WO 1999-US20594 W 19990908
WO 1999-US20944 W 19990913
WO 1999-US210 W 19990915
WO 1999-US23089 W 19991005
US 1999-162506P P 19991029
WO 1999-US28214 A 19991129
WO 1999-US28409 W 19991130
WO 1999-US28634 W 19991201
WO 1999-US28551 W 19991202
WO 1999-US28564 W 19991202
WO 1999-US28565 W 19991202
US 1999-169495P P 19991207
US 1999-170262P P 19991209
WO 1999-US30999 A 19991220
WO 1999-US31243 W 19991230
WO 1999-US31274 W 19991230
WO 2000-US277 W 20000106
US 2000-175481P P 20000111
WO 2000-US4342 W 20000218
WO 2000-US5004 W 20000224
WO 2000-US5601 W 20000301
WO 2000-US5841 W 20000302
US 2000-187202P P 20000303
US 2000-186968P P 20000306
WO 2000-US6319 W 20000310
US 2000-189320P P 20000314
US 2000-189328P P 20000314

WO 2000-US6884	W	20000315
WO 2000-US7377	W	20000320
US 2000-190828P	P	20000321
US 2000-191007P	P	20000321
US 2000-191048P	P	20000321
US 2000-191314P	P	20000321
WO 2000-US7532	W	20000321
US 2000-192655P	P	20000328
US 2000-193032P	P	20000329
US 2000-193053P	P	20000329
WO 2000-US8439	W	20000330
US 2000-194449P	P	20000404
US 2000-194647P	P	20000404
US 2000-195975P	P	20000411
US 2000-196000P	P	20000411
US 2000-196187P	P	20000411
US 2000-196690P	P	20000411
US 2000-196820P	P	20000411
US 2000-198121P	P	20000418
US 2000-198585P	P	20000418
US 2000-199397P	P	20000425
US 2000-199550P	P	20000425
US 2000-199654P	P	20000425
US 2000-201516P	P	20000503
WO 2000-US13705	W	20000517
WO 2000-US14042	W	20000522
WO 2000-US14941	W	20000530
WO 2000-US15264	W	20000602
US 2000-209832P	P	20000605
WO 2000-US20710	W	20000728
US 2000-644848	A	20000822
WO 2000-US23522	W	20000823
WO 2000-US23328	W	20000824
WO 2000-US30952	W	20001108
WO 2000-US32678	W	20001201
WO 2000-US34956	W	20001220

AB The present invention is directed to novel polypeptides and to nucleic acid mols. encoding those polypeptides. Thus, 135 cDNA sequences encoding human secreted and/or transmembrane **proteins** are identified by extracellular domain homol. screening, amylase screening, and a signal sequence algorithm to identify novel polypeptides. The **proteins** exhibit various biol. activities useful for diagnostic and therapeutic applications. Also provided herein are vectors and host cells comprising those nucleic acid sequences, chimeric polypeptide mols. comprising the polypeptides of the present invention fused to heterologous polypeptide sequences, antibodies which bind to the polypeptides of the present invention and to methods for producing the polypeptides of the present invention.

IT **252198-24-6P**
RL: ANT (Analyte); BOC (Biological occurrence); BPN (Biosynthetic preparation); BSU (Biological study, unclassified); PRP (Properties); THU (Therapeutic use); ANST (Analytical study); BIOL (Biological study); OCCU (Occurrence); PREP (Preparation); USES (Uses)
(amino acid sequence; human secreted and transmembrane polypeptides and **nucleic acids** encoding the same)

REFERENCE COUNT: 11 THERE ARE 11 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 15 OF 57 HCAPLUS COPYRIGHT 2002 ACS

ACCESSION NUMBER: 2000:829896 HCAPLUS
DOCUMENT NUMBER: 134:143447
TITLE: Structural characterisation of human proteinosis surfactant **protein** A
AUTHOR(S): Berg, T.; Leth-Larsen, R.; Holmskov, U.; Hojrup, P.
CORPORATE SOURCE: Dep. Mol. Biol., Univ. Southern Denmark, Odense Univ., Odense, DK-5230, Den.
SOURCE: Biochimica et Biophysica Acta (2000), 1543(1), 159-173
CODEN: BBACAQ; ISSN: 0006-3002
PUBLISHER: Elsevier Science B.V.
DOCUMENT TYPE: Journal
LANGUAGE: English
AB Human surfactant **protein**-A (SP-A) has been purified from a proteinosis patient and characterised by a combination of automated Edman degrdn. and mass spectrometry. The complete **protein** sequence was characterised. The major part of SP-A was shown to consist of SP-A2 **gene** product, and only a small amt. of SP-A1 **gene** product was shown to be present. A cysteine extension to the N-terminal was indicated by sequence data, but was not definitely proven. All proline residues in the Y position of Gly-X-Y in the collagen-like region were at least partially modified to hydroxy-proline, but no lysine residues were found to be modified. A complex N-linked glycosylation was found on Asn-187 showing great heterogeneity as variants from a mono-antennary to penta-antennary glycosylation with varying amts. of attached pentose were identified. The disulfide bridges in the carbohydrate recognition domain were identified to be in the 1-4, 2-3 pattern common for **collectins**. Interchain disulfide bridges were discovered between two Cys-48 residues and cysteine residues in the N-terminal region. However, the exact disulfide bridge connections within the bouquet-like ultrastructure could not be established.
REFERENCE COUNT: 42 THERE ARE 42 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 16 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 2000:602709 HCAPLUS
DOCUMENT NUMBER: 134:158282
TITLE: Identification of differentially expressed genes in epithelial stem/progenitor cells of fetal rat liver
AUTHOR(S): Petkov, Petko M.; Kim, Kwanghee; Sandhu, Jaswinder; Shafritz, David A.; Dabeva, Mariana D.
CORPORATE SOURCE: Marion Bessin Liver Research Center, Albert Einstein College of Medicine, Bronx, NY, 10461, USA
SOURCE: Genomics (2000), 68(2), 197-209
CODEN: GNMCEP; ISSN: 0888-7543
PUBLISHER: Academic Press
DOCUMENT TYPE: Journal
LANGUAGE: English
AB Differentially expressed cDNA clones from fetal rat liver were isolated using suppression subtractive hybridization, combined with an efficient screening strategy. Approx. 30,000 clones were screened, yielding 643 genes whose expression was induced, of which 201 clones were distinct and 68 represented ESTs or newly discovered genes of unknown function. Based on their expression patterns in different organs, fetal liver, liver regeneration models, and gut epithelial progenitor cell lines, the subtracted clones presented in this work were placed into four categories: (1) hepatoblast-specific genes; (2) hematopoietic cell-specific genes; (3) genes expressed in hepatoblasts, in hematopoietic cells, and at varying levels in other tissues; and (4) genes overexpressed in fetal liver, in models of activation of liver progenitor cells, and in epithelial

progenitor cell lines. Hepatoblast-specific clones and those representing genes induced during liver regeneration are under further study to define their specific function(s) in liver cell growth control and/or differentiation. (c) 2000 Academic Press.
REFERENCE COUNT: 59 THERE ARE 59 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 17 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 2000:278089 HCAPLUS
DOCUMENT NUMBER: 132:288771
TITLE: Surfactant **protein** D for the prevention and diagnosis of pulmonary emphysema
INVENTOR(S): Whitsett, Jeffrey A.
PATENT ASSIGNEE(S): Children's Hospital Medical Center, USA
SOURCE: PCT Int. Appl., 46 pp.
CODEN: PIXXD2
DOCUMENT TYPE: Patent
LANGUAGE: English
FAMILY ACC. NUM. COUNT: 1
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 2000023569	A1	20000427	WO 1999-US24675	19991020
W: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, CZ, DE, DE, DK, DK, DM, EE, EE, ES, FI, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM				
RW: GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG				
EP 1123383	A1	20010816	EP 1999-958659	19991020
R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, MC, IE, SI, LT, LV, FI, RO				
BR 9914645	A	20020205	BR 1999-14645	19991020

PRIORITY APPLN. INFO.: US 1998-104941P P 19981020
WO 1999-US24675 W 19991020

AB Surfactant **protein** D (SP-D) is a 43-kDa member of the **collectin** family of collagenous lectin domain-contg. **proteins** that is expressed in epithelial cells of the lung. The SP-D **gene** was targeted by homologous recombination in embryonic stem cells that were used to produce SP-D (-/-) mice. The SP-D (-/-) deficiency caused inflammation, increased oxidant prodn. by isolated alveolar macrophages, abnormal surfactant structure and levels, and decreased SP-A expression. Therefore, disclosed is the SP-D (-/-) mouse as an excellent model for emphysema. Also included are models for testing emphysema therapies in the mouse model, methods for using SP-D **protein** or **DNA** as a treatment for emphysema and pulmonary infections, and diagnosis.
REFERENCE COUNT: 7 THERE ARE 7 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 18 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 2000:222973 HCAPLUS
DOCUMENT NUMBER: 133:3124
TITLE: DMBT1 encodes a **protein** involved in the immune defense and in epithelial differentiation and

is highly unstable in cancer
AUTHOR(S): Mollenhauer, Jan; Herbertz, Stephan; Holmskov, Uffe; Tolnay, Markus; Krebs, Inge; Merlo, Adrian; Schroder, Henrik Daa; Maier, Daniel; Breitling, Frank; Wiemann, Stefan; Grone, Hermann-Josef; Poustka, Annemarie
CORPORATE SOURCE: Department of Molecular Genome Analysis, Deutsches Krebsforschungszentrum, Heidelberg, 69120, Germany
SOURCE: Cancer Research (2000), 60(6), 1704-1710
CODEN: CNREA8; ISSN: 0008-5472
PUBLISHER: AACR Subscription Office
DOCUMENT TYPE: Journal
LANGUAGE: English
AB The **gene** deleted in malignant brain tumors 1 (DMBT1) has been proposed as a candidate tumor suppressor for brain, gastrointestinal, and lung cancer. It codes for a **protein** of unknown function belonging to the superfamily of scavenger receptor cysteine-rich **proteins**. The authors aimed at getting insights into the functions of DMBT1 by expression analyses and studies with a monoclonal antibody against the **protein**. The DMBT1 mRNA is expressed throughout the immune system, and Western blot studies demonstrated that isoforms of DMBT1 are identical to the **collectin**-binding **protein** gp-340, a glycoprotein that is involved in the respiratory immune defense. Immunohistochem. analyses revealed that DMBT1 is produced by both tumor-assocd. macrophages and tumor cells and that it is deregulated in glioblastoma multiforme in comparison to normal brain tissue. The data further suggest that the **proteins** CRP-ductin and hensin, both of which have been implicated in epithelial differentiation, are the DMBT1 orthologs in mice and rabbits, resp. These findings and the spatial and temporal distribution of DMBT1 in fetal and adult epithelia suggest that DMBT1 further plays a role in epithelial development. Rearrangements of DMBT1 were found in 16 of 18 tumor cell lines, and hemizygous deletions were obsd. in a subset of normal individuals, indicating that the alterations in tumors may be a result of both pre-existing deletions uncovered by a loss of heterozygosity and secondary changes acquired during tumorigenesis. Thus, DMBT1 is a **gene** that is highly unstable in cancer and encodes for a **protein** with at least two different functions, one in the immune defense and a second one in epithelial differentiation.
REFERENCE COUNT: 28 THERE ARE 28 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 19 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 2000:216336 HCAPLUS
DOCUMENT NUMBER: 132:235583
TITLE: **Collectin** family as host-defense lectins
AUTHOR(S): Wakamiya, Nobutaka; Suzuki, Yasuhiko
CORPORATE SOURCE: Res. Inst. for Microb. Dis., Osaka Univ., Yamada-oka, Suita, Osaka, 565-0871, Japan
SOURCE: Tanpakushitsu Kakusan Koso (2000), 45(5), 655-663
CODEN: TAKKAJ; ISSN: 0039-9450
PUBLISHER: Kyoritsu Shuppan
DOCUMENT TYPE: Journal; General Review
LANGUAGE: Japanese
AB A review with 47 refs., on the structure and **genes** of **collectins**, physiol. functions of **collectins**, mannan-binding lectin deficiency, mechanisms regulating blood levels of **collectins**, role of collagen-like domain, structure of carbohydrate recognition domains in relation to sugar specificity of **collectins**, roles of **collectins** in host defense against

infections, and characterization of a newly cloned **collectin** CL-L1.

L6 ANSWER 20 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 2000:145014 HCAPLUS
DOCUMENT NUMBER: 132:204040
TITLE: Cloning of cDNA for novel human **collectin** for developing antibacterial and antiviral drugs
INVENTOR(S): Wakamiya, Nobutaka
PATENT ASSIGNEE(S): Fuso Pharmaceutical Industries, Ltd., Japan
SOURCE: PCT Int. Appl., 106 pp.
CODEN: PIXXD2
DOCUMENT TYPE: Patent
LANGUAGE: Japanese
FAMILY ACC. NUM. COUNT: 1
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 2000011161	A1	20000302	WO 1999-JP4552	19990824
W: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM				
RW: GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG				
AU 9953056	A1	20000314	AU 1999-53056	19990824
EP 1108786	A1	20010620	EP 1999-938607	19990824
R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, SI, LT, LV, FI, RO				

PRIORITY APPLN. INFO.: JP 1998-237611 A 19980824
WO 1999-JP4552 W 19990824

AB The cDNA encoding a novel **collectin** is isolated from a human placenta cDNA library by using the screening probes prepd. from a human fetus clone (I.M.A.G.E. Consortium Clone ID 34472). The novel **collectin** is comprised of 342 amino acids that contains a Ca2+-dependent carbohydrate recognition domain (CRD) and a collagen-like domain. Human cells contain a single copy of the **collectin gene**. Also described are monoclonal antibodies to the **collectin** and use of immunoassay, oligonucleotide probes derived from the cDNA, transgenic mice expressing the **collectin**, the **collectin gene** knockout mice, etc. Amino acid sequences deduced from other open reading frames in the cDNA sequence are also shown. The novel **collectin** can be used for developing antibacterial and antiviral drugs.

IT **260234-88-6, Collectin** (human) **260234-89-7,** 24-342-**Collectin** (human)
RL: BSU (Biological study, unclassified); PRP (Properties); THU (Therapeutic use); BIOL (Biological study); USES (Uses)
(amino acid sequence; cloning of cDNA for novel human **collectin** for developing antibacterial and antiviral drugs)

IT **260234-86-4, DNA** (human **collectin** cDNA plus flanks) **260234-94-4, DNA** (human **collectin** cDNA) **260234-95-5 260234-96-6 260234-97-7 260234-98-8 260234-99-9 260235-00-5 260235-01-6 260235-02-7, DNA** (human

collectin cDNA 3'-flank)
RL: BSU (Biological study, unclassified); BUU (Biological use, unclassified); PRP (Properties); BIOL (Biological study); USES (Uses)
(nucleotide sequence; cloning of cDNA for novel human **collectin** for developing antibacterial and antiviral drugs)
REFERENCE COUNT: 5 THERE ARE 5 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 21 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 2000:82595 HCAPLUS
DOCUMENT NUMBER: 132:220182
TITLE: Porcine lung surfactant **protein** D: complementary DNA cloning, chromosomal localization, and tissue distribution
AUTHOR(S): Van Eijk, Martin; Haagsman, Henk P.; Skinner, Thomas; Archibold, Alan; Reid, Kenneth B. M.; Lawson, Peter R.
CORPORATE SOURCE: Laboratory of Veterinary Biochemistry and Graduate School of Animal Health, Utrecht University, Utrecht, 3508 TD, Neth.
SOURCE: Journal of Immunology (2000), 164(3), 1442-1450
CODEN: JOIMA3; ISSN: 0022-1767
PUBLISHER: American Association of Immunologists
DOCUMENT TYPE: Journal
LANGUAGE: English
AB Porcine organs and lung surfactant have medically important applications in both xenotransplantation and therapy. The authors have started to characterize porcine lung surfactant by cloning the cDNA of porcine surfactant **protein** D (SP-D). SP-D and SP-A are important mediators in innate immune defense for the lung and possibly other mucosal surfaces. Porcine SP-D will also be an important reagent for use in existing porcine animal models for human lung infections. The complete cDNA sequence of porcine SP-D, including the 5' and 3' untranslated regions, was detd. from two overlapping bacteriophage clones and by PCR cloning. Three unique features were revealed from the porcine sequence in comparison to SP-D from other previously characterized species, making porcine SP-D an intriguing species addn. to the SP-D/**collectin** family. The collagen region contains an extra cysteine residue, which may have important structural consequences. The other two differences, a potential glycosylation site and an insertion of three amino acids, lie in the loop regions of the carbohydrate recognition domain, close to the carbohydrate binding region and thus may have functional implications. These variations were ruled out as polymorphisms or mutations by confirming the sequence at the genomic level in four different pig breeds. Porcine SP-D was shown to localize primarily to the lung and with less abundance to the duodenum, jejunum, and ileum. The **genes** for SP-D and SP-A were also shown to colocalize to a region of porcine chromosome 14 that is syntenic with the human and murine **collectin** loci.
REFERENCE COUNT: 59 THERE ARE 59 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 22 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 2000:73320 HCAPLUS
DOCUMENT NUMBER: 132:217921
TITLE: GATA-6 activates transcription of surfactant **protein** A
AUTHOR(S): Bruno, Michael D.; Korfhagen, Thomas R.; Liu, Cong; Morrisey, Edward E.; Whitsett, Jeffrey A.
CORPORATE SOURCE: Division of Pulmonary Biology, Children's Hospital

Medical Center, Cincinnati, OH, 45229-3039, USA
SOURCE: Journal of Biological Chemistry (2000), 275(2), 1043-1049
CODEN: JBCHA3; ISSN: 0021-9258
PUBLISHER: American Society for Biochemistry and Molecular Biology
DOCUMENT TYPE: Journal
LANGUAGE: English
AB Surfactant **protein** A (SP-A) is a member of the **collectin** family of innate host defense mols. expressed primarily in respiratory epithelial cells of the lung. SP-A concns. are influenced by both cell-specific and ubiquitous nuclear **proteins** that regulate SP-A **gene** transcription in a cell-selective and temporally regulated manner. In this work, a consensus GATA-binding site (GBS) was identified at positions -69 to -64 of the mouse SP-A **gene**. The transcriptional activity of wild-type SP-A reporter constructs in HeLa cells was increased 5-10-fold when cotransfected with a GATA-6 expression plasmid. Deletion of the GBS completely blocked transactivation by GATA-6. Transfection of a construct expressing GATA-6-engrailed fusion **protein** inhibited basal expression of the SP-A/chloramphenicol acetyltransferase construct in MLE-15 cells. Nuclear ext. **proteins** from MLE-15 cells bound to the GBS in the mouse SP-A **gene**, and a supershifted band was detected with a GATA-6-specific antibody. Transactivation of the wild-type SP-A constructs by GATA-6 increased transcriptional activity 7-10-fold, whereas thyroid transcription factor-1 (TTF-1) increased the activity of these constructs 12-18-fold. The effects of cotransactivating with both GATA-6 and TTF-1 expression constructs were additive. However, mutation of the TTF-1-binding sites alone or in combination decreased GATA-6 transactivation. Likewise, mutation of the GBS blocked TTF-1 activation of the SP-A promoter. In situ hybridization demonstrated GATA-6 mRNA in the peripheral epithelial cells of fetal mouse lung, consistent with the sites of SP-A expression. GATA-6 is expressed in respiratory epithelial cells and binds to a cis-acting element in the SP-A **gene** promoter, activating the transcriptional activity of the **gene**.
REFERENCE COUNT: 32 THERE ARE 32 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 23 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 1999:784264 HCAPLUS
DOCUMENT NUMBER: 132:31785
TITLE: Nucleic acids encoding membrane-bound **proteins** from human
INVENTOR(S): Baker, Kevin; Chen, Jian; Goddard, Audrey; Gurney, Austin L.; Smith, Victoria; Watanabe, Colin K.; Wood, William I.; Yuan, Jean
PATENT ASSIGNEE(S): Genentech, Inc., USA
SOURCE: PCT Int. Appl., 822 pp.
CODEN: PIXXD2
DOCUMENT TYPE: Patent
LANGUAGE: English
FAMILY ACC. NUM. COUNT: 71
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 9963088	A2	19991209	WO 1999-US12252	19990602
WO 9963088	A3	20010329		

W: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ,

DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS,
JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK,
MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ,
TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ,
MD, RU, TJ, TM

RW: GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW, AT, BE, CH, CY, DE, DK,
ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG,
CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG

CA 2328895 AA 19991209 CA 1999-2328895 19990602
AU 9943286 A1 19991220 AU 1999-43286 . 19990602
WO 2000032221 A2 20000608 WO 1999-US28313 19991130
WO 2000032221 A3 20001123

W: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU,
CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL,
IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD,
MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK,
SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ,
BY, KG, KZ, MD, RU, TJ, TM

RW: GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE,
DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF,
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG

EP 1135485 A2 20010926 EP 1999-960624 19991130

R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT,
IE, SI, LT, LV, FI, RO

WO 2000036102 A2 20000622 WO 1999-US28634 19991201
WO 2000036102 A3 20001109

W: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU,
CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL,
IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA,
MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI,
SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM,
AZ, BY, KG, KZ, MD, RU, TJ, TM

RW: GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE,
DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF,
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG

EP 1141285 A2 20011010 EP 1999-962992 19991201

R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT,
IE, SI, LT, LV, FI, RO

WO 2000037640 A2 20000629 WO 1999-US30095 19991216
WO 2000037640 A3 20001228

W: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU,
CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL,
IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD,
MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK,
SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY,
KG, KZ, MD, RU, TJ, TM

RW: GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE,
DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF,
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG

WO 2000075316 A1 20001214 WO 1999-US30911 19991220

W: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU,
CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL,
IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD,
MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK,
SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY,
KG, KZ, MD, RU, TJ, TM

RW: GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE,
DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF,
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG

WO 2000053753 A2 20000914 WO 2000-US219 20000105
WO 2000053753 A3 20010419
W: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM
RW: GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG
WO 2000053755 A2 20000914 WO 2000-US376 20000106
WO 2000053755 A3 20010208
W: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM
RW: GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG
WO 2001053486 A1 20010726 WO 2000-US3565 20000211
W: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM
RW: GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG
EP 1173563 A1 20020123 EP 2000-907270 20000211
R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, SI, LT, LV, FI, RO
WO 2000053757 A2 20000914 WO 2000-US5004 20000224
WO 2000053757 A3 20011101
W: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM
RW: GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG
WO 2000075327 A1 20001214 WO 2000-US4914 20000224
W: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM
RW: GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG
EP 1159419 A1 20011205 EP 2000-912015 20000224
R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, SI, LT, LV, FI, RO

WO 2000053758 A2 20000914 WO 2000-US5841 20000302
WO 2000053758 A3 20010208
W: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM
RW: GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG

WO 2000073454 A1 20001207 WO 2000-US8439 20000330
W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM
RW: GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG

EP 1210418 A1 20020605 EP 2000-916675 20000330
R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, SI, LT, LV, FI, RO, MK, CY, AL

WO 2000073445 A2 20001207 WO 2000-US13705 20000517
WO 2000073445 A3 20010301
W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM
RW: GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG

EP 1212417 A2 20020612 EP 2000-939307 20000517
R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, SI, LT, LV, FI, RO, MK, CY, AL

WO 2000073348 A2 20001207 WO 2000-US14941 20000530
WO 2000073348 A3 20010705
W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM
RW: GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG

EP 1185648 A2 20020313 EP 2000-941164 20000530
R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, SI, LT, LV, FI, RO

WO 2000073452 A2 20001207 WO 2000-US15264 20000602
WO 2000073452 A3 20010628
W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU,

ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM
RW: GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY,
DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ,
CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG

PRIORITY APPLN. INFO.:

US	1998-87607P	P	19980602
US	1998-87609P	P	19980602
US	1998-87759P	P	19980602
US	1998-87827P	P	19980603
US	1998-88021P	P	19980604
US	1998-88025P	P	19980604
US	1998-88028P	P	19980604
US	1998-88029P	P	19980604
US	1998-88030P	P	19980604
US	1998-88033P	P	19980604
US	1998-88326P	P	19980604
US	1998-88167P	P	19980605
US	1998-88202P	P	19980605
US	1998-88212P	P	19980605
US	1998-88217P	P	19980605
US	1998-88655P	P	19980609
US	1998-88722P	P	19980610
US	1998-88730P	P	19980610
US	1998-88734P	P	19980610
US	1998-88738P	P	19980610
US	1998-88740P	P	19980610
US	1998-88741P	P	19980610
US	1998-88742P	P	19980610
US	1998-88810P	P	19980610
US	1998-88811P	P	19980610
US	1998-88824P	P	19980610
US	1998-88825P	P	19980610
US	1998-88826P	P	19980610
US	1998-88858P	P	19980611
US	1998-88861P	P	19980611
US	1998-88863P	P	19980611
US	1998-88876P	P	19980611
US	1998-89090P	P	19980612
US	1998-89105P	P	19980612
US	1998-89440P	P	19980616
US	1998-89512P	P	19980616
US	1998-89514P	P	19980616
US	1998-89532P	P	19980617
US	1998-89538P	P	19980617
US	1998-89598P	P	19980617
WO	1998-US25108	W	19981201
US	1998-112850P	P	19981216
US	1998-112851P	P	19981216
US	1998-113145P	P	19981216
US	1998-113296P	P	19981222
US	1998-113511P	P	19981222
US	1999-115554P	P	19990112
US	1999-115558P	P	19990112
US	1999-115565P	P	19990112
US	1999-115733P	P	19990112
US	1999-119341P	P	19990209
US	1999-119537P	P	19990210
US	1999-119965P	P	19990212
WO	1999-US5028	W	19990308
US	1999-123618P	P	19990310

US	1999-123972P	P	19990311
US	1999-123957P	P	19990312
US	1999-125775P	P	19990323
US	1999-128849P	P	19990412
WO	1999-US8615	W	19990420
US	1999-131445P	P	19990428
US	1999-132371P	P	19990504
US	1999-133459P	P	19990511
US	1999-134287P	P	19990514
WO	1999-US12252	W	19990602
US	1999-140650P	P	19990622
US	1999-140653P	P	19990622
US	1999-141037P	P	19990623
US	1999-143048P	P	19990707
US	1999-144732P	P	19990720
US	1999-144758P	P	19990720
US	1999-145698P	P	19990726
US	1999-146222P	P	19990728
US	1999-149395P	P	19990817
US	1999-149396P	P	19990817
US	1999-151689P	P	19990831
WO	1999-US20111	W	19990901
WO	1999-US20594	W	19990908
WO	1999-US20944	W	19990913
WO	1999-US21090	W	19990915
WO	1999-US21547	W	19990915
WO	1999-US23089	W	19991005
US	1999-158663P	P	19991008
US	1999-162506P	P	19991029
WO	1999-US28214		19991129
WO	1999-US28313	W	19991130
WO	1999-US28409	W	19991130
WO	1999-US28301	W	19991201
WO	1999-US28634	W	19991201
WO	1999-US28551		19991202
WO	1999-US28564	W	19991202
WO	1999-US28565	W	19991202
US	1999-170262P	P	19991209
WO	1999-US30095		19991216
WO	1999-US30911	W	19991220
WO	1999-US30999		19991220
WO	1999-US31274		19991230
WO	2000-US219	W	20000105
WO	2000-US277		20000106
WO	2000-US376	A	20000106
WO	2000-US3565	W	20000211
WO	2000-US4341	A	20000218
WO	2000-US4342	A	20000218
WO	2000-US4414	A	20000222
WO	2000-US4914	A	20000224
WO	2000-US5004	W	20000224
WO	2000-US5841	W	20000302
US	2000-187202P	P	20000303
WO	2000-US6319	A	20000310
WO	2000-US6884	W	20000315
WO	2000-US7377	W	20000320
WO	2000-US7532	A	20000321
WO	2000-US8439	W	20000330
WO	2000-US13705	W	20000517

WO 2000-US14941 W 20000530

AB The present invention is directed to 135 polypeptides and to nucleic acid mols. encoding those polypeptides. The extracellular domain sequences (including the secretion signal sequence, if any) from about 950 known secreted **proteins** from the Swiss-Prot public database were used to search EST (expressed sequence tag) databases, and this homol. screen used to assemble consensus DNA sequences relative to other identified EST sequences. Based upon the consensus sequences obtained, oligonucleotides were than synthesized and used to identify by PCR a cDNA library that contained the sequences of interest and for use as probes to isolate clones of full-length coding sequences for the PRO polypeptides. Also provided herein are vectors and host cells comprising those nucleic acid sequences, chimeric polypeptide mols. comprising the polypeptides of the present invention fused to heterologous polypeptide sequences, antibodies which bind to the polypeptides of the present invention and to methods for producing the polypeptides of the present invention. This invention is particularly useful for screening compds. by using PRO polypeptides or binding fragment thereof in any of a variety of drug screening techniques.

IT **252198-24-6P**
RL: BOC (Biological occurrence); BPN (Biosynthetic preparation); BSU (Biological study, unclassified); PRP (Properties); THU (Therapeutic use); BIOL (Biological study); OCCU (Occurrence); PREP (Preparation); USES (Uses)
(amino acid sequence; **nucleic acids** encoding membrane-bound **proteins** from human)

L6 ANSWER 24 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 1999:622680 HCAPLUS
DOCUMENT NUMBER: 131:335244
TITLE: SP-A enhances viral clearance and inhibits inflammation after pulmonary adenoviral infection
AUTHOR(S): Harrod, Kevin S.; Trapnell, Bruce C.; Otake, Kazuhisa; Korfhagen, Thomas R.; Whitsett, Jeffrey A.
CORPORATE SOURCE: Division of Neonatology and Pulmonary Biology, Children's Hospital Medical Center, Cincinnati, OH, 45229, USA
SOURCE: American Journal of Physiology (1999), 277(3, Pt. 1), L580-L588
CODEN: AJPHAP; ISSN: 0002-9513
PUBLISHER: American Physiological Society
DOCUMENT TYPE: Journal
LANGUAGE: English

AB Surfactant **protein** A (SP-A) is a member of the **collectin** family of host defense mols. expressed primarily in the epithelial cells of the lung. To det. the role of SP-A in pulmonary adenoviral infection, SP-A-deficient (SP-A -/-) mice were intratracheally infected with a replication-deficient recombinant adenovirus, Av1Luc1. Lung inflammation was markedly increased in SP-A -/- compared with SP-A +/+ mice and was assocd. with increased hemorrhage and epithelial cell injury. Polymorphonuclear cells in bronchoalveolar lavage fluid (BALF) were increased in SP-A -/- mice after administration of adenovirus. Coadministration of adenovirus and purified human SP-A ameliorated adenoviral-induced lung inflammation in SP-A -/- mice. Concns. of tumor necrosis factor-.alpha. (TNF-.alpha.), interleukin (IL)-6, and IL-1.beta. were increased in BALF of SP-A -/- mice. Likewise, TNF-.alpha., IL-6, macrophage inflammatory **protein** (MIP)-1.alpha., monocyte chemotactic **protein**-1, and MIP-2 mRNAs were increased in lung homogenates from SP-A -/- mice 6 and 24 h after viral administration. Clearance of adenoviral **DNA** from the lung and uptake of

fluorescent-labeled adenovirus by alveolar macrophages were decreased in SP-A -/- mice. SP-A enhances viral clearance and inhibits lung inflammation during pulmonary adenoviral infection, providing support for the importance of SP-A in antiviral host defense.
REFERENCE COUNT: 41 THERE ARE 41 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 25 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 1999:547704 HCAPLUS
DOCUMENT NUMBER: 131:282260
TITLE: Characterization of the mouse **collectin gene** locus
AUTHOR(S): Akiyama, Jennifer; Volik, Stanislav V.; Plajzer-Frick, Ingrid; Prince, Amy; Sago, Haruhiko; Weier, Heinz-Ulrich G.; Vanderbilt, Jeff N.; Hawgood, Sam; Poulain, Francis R.
CORPORATE SOURCE: Cardiovascular Research Institute and Department of Pediatrics, University of California San Francisco, San Francisco, CA, 94118-1245, USA
SOURCE: American Journal of Respiratory Cell and Molecular Biology (1999), 21(2), 193-199
CODEN: AJRBEL; ISSN: 1044-1549
PUBLISHER: American Lung Association
DOCUMENT TYPE: Journal
LANGUAGE: English
AB Three of the four known mouse **collectin genes** have been mapped to chromosome 14. To further characterize the spatial relation of these **genes**, a bacterial artificial chromosome (BAC) library of mouse chromosome 14 was screened using mouse surfactant **protein** (SP)-A and -D complementary **DNAs** (cDNAs). One large clone hybridized to both SP-A and SP-D cDNAs and was found by polymerase chain reaction (PCR) to contain sequences from one of the mouse mannose-binding lectin **genes** (Mbl1). The authors used Southern mapping and subcloning of overlapping restriction fragments to characterize the **gene** locus. Mapping was confirmed by fluorescent in situ hybridization of fiber-stretched BAC **DNA** and by Southern hybridization of restriction endonuclease-digested and PCR-amplified genomic **DNA**. The authors found that the SP-A, Mbl1, and SP-D **genes** reside contiguously within a 55-kb region. The SP-A and Mbl1 **genes** are in the same 5' to 3' orientation and 16 kb apart. The SP-D **gene** is in the opposite orientation to the two other **collectin genes**, 13 kb away from the 3' end of the Mbl1 **gene**. The mouse SP-D **gene** had not previously been characterized. The authors found its size (13 kb) and organization to be similar to that of human SP-D. Exon I is untranslated. The second exon is a hybrid exon that contains signal for initiation of translation, signal peptide, N-terminal domain, and the first seven collagen triplets of the collagen-like domain of the **protein**. Four short exons (III through VI) encode the collagen-like domain of the **protein**, and exons VII and VIII the linking and the carbohydrate-recognition domains, resp.
REFERENCE COUNT: 42 THERE ARE 42 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 26 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 1999:480717 HCAPLUS
DOCUMENT NUMBER: 131:156226
TITLE: Cloning of cDNA for novel human **collectin**
INVENTOR(S): Wakamiya, Nobutaka

PATENT ASSIGNEE(S): Fuso Pharmaceutical Industries, Ltd., Japan
SOURCE: Jpn. Kokai Tokkyo Koho, 18 pp.
CODEN: JKXXAF
DOCUMENT TYPE: Patent
LANGUAGE: Japanese
FAMILY ACC. NUM. COUNT: 1
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
JP 11206377	A2	19990803	JP 1998-11281	19980123
CA 2319084	AA	19990729	CA 1998-2319084	19980724
WO 9937767	A1	19990729	WO 1998-JP3328	19980724

W: CA, US

PRIORITY APPLN. INFO.: JP 1998-11281 A 19980123
WO 1998-JP3328 W 19980724

AB A novel human **collectin** is identified and its encoding cDNA sequence is isolated by screening a human liver cDNA library using the primers/probes derived from GenBank No. R29493 that contains a consensus sequence among human **collectins** such MBP, SP-A and SP-D. The **collectin** is characterized as having (1) Ca2+-dependent carbohydrate recognition domain (CRD); (2) a neck domain; (3) a collagen-like domain; and (4) a cysteine-contg. N-terminus. The **collectin** may be used for developing antiviral agents.

IT **235094-70-9**
RL: BSU (Biological study, unclassified); PRP (Properties); BIOL (Biological study)
(amino acid sequence; cloning of cDNA for novel human **collectin**)

IT **236742-12-4**
RL: BOC (Biological occurrence); BSU (Biological study, unclassified); PRP (Properties); BIOL (Biological study); OCCU (Occurrence)
(nucleotide sequence; cloning of cDNA for novel human **collectin**)

L6 ANSWER 27 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 1999:464352 HCAPLUS
DOCUMENT NUMBER: 131:101254
TITLE: Vaccine comprising nucleic acids encoding a fusion **protein** of an antiten and an APC-binding domain of an opsonin and methods of modulating immune responses
INVENTOR(S): Segal, Andrew
PATENT ASSIGNEE(S): Genitrix, LLC, USA
SOURCE: PCT Int. Appl., 102 pp.
CODEN: PIXXD2
DOCUMENT TYPE: Patent
LANGUAGE: English
FAMILY ACC. NUM. COUNT: 3
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 9936507	A1	19990722	WO 1999-US894	19990115

W: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, AM, AZ, BY, KG, KZ, MD, RU,

TJ, TM
RW: GH, GM, KE, LS, MW, SD, SZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG

US 6224870	B1	20010501	US 1998-7711	19980115
AU 9922301	A1	19990802	AU 1999-22301	19990115

PRIORITY APPLN. INFO.:		
US 1998-7711	A	19980115
US 1997-788143	B2	19970124
WO 1999-US894	W	19990115

AB Methods of modulating an immune response are disclosed wherein a fusion **protein** comprising an antigen and an APC (antigen presenting cells) binding domain of an opsonin or a nucleic acid encoding a fusion **protein** comprising an antigen and an APC binding domain of an opsonin is administered as an immunogen. A fusion gene encoding for the antigen hen egg lysozyme and for the .alpha. chain of the opsonin murine C3b or murine mannose binding **protein** A (MBP) was generated and administered to mice. Either C3b .alpha. chain or MBP was able to markedly attenuate antibody response to hen egg lysosome.

REFERENCE COUNT: 7 THERE ARE 7 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 28 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 1999:424919 HCAPLUS
DOCUMENT NUMBER: 131:211955
TITLE: Surfactant **protein** A and D expression in the porcine Eustachian tube
AUTHOR(S): Paananen, Reija; Glumoff, Virpi; Hallman, Mikko
CORPORATE SOURCE: PL 5000, Kajaanintie 52A, Biocenter Oulu and Department of Pediatrics, University of Oulu, Oulu, 90220, Finland
SOURCE: FEBS Letters (1999), 452(3), 141-144
CODEN: FEBLAL; ISSN: 0014-5793
PUBLISHER: Elsevier Science B.V.
DOCUMENT TYPE: Journal
LANGUAGE: English

AB Surfactant **proteins** A and D are **collectins** which are considered to play an important role in the innate immunity of lungs. Our aim was to investigate whether surfactant **protein** A or D is expressed in the porcine Eustachian tube originating from the upper airways. Both surfactant **proteins** A and D were present in the epithelial cells of the Eustachian tube, as shown by strong immunostaining. Using RT-PCR and Northern hybridization, these **collectins** were detected in the Eustachian tube. The present study is the first report demonstrating surfactant **protein** **gene** expression in the Eustachian tube. Surfactant **proteins** A and D may be important in the antibody-independent protection of the middle ear.

REFERENCE COUNT: 30 THERE ARE 30 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 29 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 1999:336101 HCAPLUS
DOCUMENT NUMBER: 131:140286
TITLE: Genomic organization of the mouse gene for lung surfactant **protein** D
AUTHOR(S): Lawson, Peter R.; Perkins, Vivienne C.; Holmskov, Uffe; Reid, Kenneth B. M.
CORPORATE SOURCE: MRC Immunochemistry Unit, Department of Biochemistry, Oxford University, Oxford, OX1 3QU, UK

SOURCE: American Journal of Respiratory Cell and Molecular Biology (1999), 20(5), 953-963
CODEN: AJRBEL; ISSN: 1044-1549
PUBLISHER: American Lung Association
DOCUMENT TYPE: Journal
LANGUAGE: English

AB Lung surfactant **protein** (SP)-D belongs to the family of sol. collagenous C-type lectins, named **collectins**. SP-D participates in the local innate immune defense of the lung, eliciting various effector functions by acting as a pattern recognition receptor for the carbohydrate structures on inhaled microorganisms and particulate matter. This work describes the isolation and characterization of the mouse SP-D **gene** (Sftpd), which spans 8 exons over 14 kb of sequence and shows an overall organization similar to other **collectin genes**. The complete 5' untranslated region of the mRNA, absent from the published complementary **DNA** for mouse SP-D, was also cloned and is shown to be encoded by a single exon. Anal. of 3.5 kb of 5' flanking nucleotide sequence for Sftpd is described and reveals positional conservation of a no. of transcription factor binding sites on comparison of Sftpd with the human SP-D **gene** and the bovine conglutinin **gene**. In addn., a single copy SP-D-like **gene** has been shown to be present in mammals, birds, and amphibians but is absent in fish. An atypical, rodent-specific, long terminal repeat of retroviral origin contg. a minisatellite that has become inserted in Sftpd is described. Three new polymorphic microsatellites are also described, one of which is just 160 base pairs upstream of Sftpd. This microsatellite was used to map the **gene** to the central region of chromosome 14; fine-scale mapping indicates that it lies in a 5.64-centimorgan area between D14Mit45 and D14Mit60. This will allow the easy identification of the **collectin gene** cluster and aid in the construction of a phys. map over this region.

REFERENCE COUNT: 49 THERE ARE 49 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 30 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 1999:317813 HCAPLUS
DOCUMENT NUMBER: 131:141029
TITLE: Molecular cloning of a novel human **collectin** from liver (CL-L1)
AUTHOR(S): Ohtani, Katsuki; Suzuki, Yasuhiko; Eda, Souji; Kawai, Takao; Kase, Tetsuo; Yamazaki, Hiroshi; Shimada, Tsutomu; Keshi, Hiroyuki; Sakai, Yoshinori; Fukuoh, Atsushi; Sakamoto, Takashi; Wakamiya, Nobutaka
CORPORATE SOURCE: Department of Pathology, Osaka Prefectural Institute of Public Health, Higashinari, Osaka, 537, Japan
SOURCE: Journal of Biological Chemistry (1999), 274(19), 13681-13689
CODEN: JBCHA3; ISSN: 0021-9258
PUBLISHER: American Society for Biochemistry and Molecular Biology
DOCUMENT TYPE: Journal
LANGUAGE: English

AB **Collectins** are a C-lectin family with collagen-like sequences and carbohydrate recognition domains. These **proteins** can bind to carbohydrate antigens of microorganisms and inhibit their infection by direct neutralization and agglutination, the activation of complement through the lectin pathway, and opsonization by **collectin** receptors. Here we report the cloning of a cDNA encoding human **collectin** from liver (CL-L1 (**collectin** liver 1)) that

has typical **collectin** structural characteristics, consisting of an N-terminal cysteine-rich domain, a collagen-like domain, a neck domain, and a carbohydrate recognition domain. The cDNA has an insert of 831 base pairs coding for a **protein** of 277 amino acid residues. The deduced amino acid sequence shows that this **collectin** has a unique repeat of four lysine residues in its C-terminal area. Northern blot, Western blot, and RT-PCR analyses showed that CL-L1 is present mainly in liver as a cytosolic **protein** and at low levels in placenta. More sensitive analyses by RT-PCR showed that most tissues (except skeletal muscle) have CL-L1 mRNA. Zoo-blot anal. indicated that CL-L1 is limited to mammals and birds. A chromosomal localization study indicated that the CL-L1 **gene** localizes to chromosome 8q23-q24.1, different from chromosome 10 of other human **collectin genes**. Expression studies of fusion **proteins** lacking the collagen and N-terminal domains produced in Escherichia coli affirmed that CL-L1 binds mannose weakly. CL-L1 and recombinant CL-L1 fusion **proteins** do not bind to mannan columns. Anal. of the phylogenetic tree of CL-L1 and other **collectins** indicated that CL-L1 belongs to a fourth subfamily of **collectins** following the mannan-binding **protein**, surfactant **protein** A, and surfactant **protein** D subfamilies including bovine conglutinin and **collectin**-43 (CL-43). These findings indicate that CL-L1 may be involved in different biol. functions.

IT **235094-70-9**
RL: BSU (Biological study, unclassified); PRP (Properties); BIOL (Biological study)
(amino acid sequence; mol. cloning of novel human **collectin** from liver (CL-L1))

IT **226543-52-8**, GenBank AB002631
RL: BSU (Biological study, unclassified); PRP (Properties); BIOL (Biological study)
(nucleotide sequence; mol. cloning of novel human **collectin** from liver (CL-L1))

REFERENCE COUNT: 38 THERE ARE 38 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 31 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 1999:132403 HCAPLUS
DOCUMENT NUMBER: 130:295176
TITLE: Lung surfactant **proteins** involved in innate immunity
AUTHOR(S): Eggleton, Paul; Reid, Kenneth B. M.
CORPORATE SOURCE: MRC Immunochemistry Unit, Department of Biochemistry, University of Oxford, Oxford, OX1 3QU, UK
SOURCE: Current Opinion in Immunology (1999), 11(1), 28-33
CODEN: COPIEL; ISSN: 0952-7915
PUBLISHER: Current Biology Publications
DOCUMENT TYPE: Journal; General Review
LANGUAGE: English

AB A review with 36 refs. The two lung surfactant **collectins**, surfactant **protein** (SP)-A and SP-D, are involved in host defense against infectious and allergenic agents via enhancement of killing and clearance by macrophages and neutrophils. Recent **gene**-knockout, **protein** engineering and physiol. studies have emphasized the roles that SP-A and SP-D play in acute inflammatory responses.

REFERENCE COUNT: 36 THERE ARE 36 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 32 OF 57 HCAPLUS COPYRIGHT 2002 ACS

ACCESSION NUMBER: 1999:130595 HCAPLUS
DOCUMENT NUMBER: 130:195768
TITLE: C1q and **collectin** receptor
INVENTOR(S): Schwaeble, Wilhelm
PATENT ASSIGNEE(S): University of Leicester, UK
SOURCE: PCT Int. Appl., 31 pp.
CODEN: PIXXD2
DOCUMENT TYPE: Patent
LANGUAGE: English
FAMILY ACC. NUM. COUNT: 1
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 9907406	A1	19990218	WO 1998-GB2430	19980812
W: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM				
RW: GH, GM, KE, LS, MW, SD, SZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG				
AU 9887406	A1	19990301	AU 1998-87406	19980812
EP 1003544	A1	20000531	EP 1998-938805	19980812
R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, FI				
JP 2001513513	T2	20010904	JP 2000-506995	19980812

PRIORITY APPLN. INFO.: GB 1997-16998 A 19970812
WO 1998-GB2430 W 19980812

AB The present invention concerns novel uses of the cC1q Receptor (cC1qR) binding domain and inhibitors thereof. The C1q receptor inhibitors are useful for inhibition of CUB (complement ubiquitin) domain functionality and for treatment of complement activation involved in inflammation, myocardial infarction, brain ischemia, gut ischemia, rheumatoid arthritis, systemic lupus erythematosus, burns, immune complex nephritis, or amyloid plaques in Alzheimer's disease.

REFERENCE COUNT: 7 THERE ARE 7 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 33 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 1999:74017 HCAPLUS
DOCUMENT NUMBER: 130:263032
TITLE: Functional characterization of the bovine conglutinin promoter: presence of a novel element for transcriptional regulation of a C-type mammalian lectin containing a collagen-like domain
AUTHOR(S): Kawasaki, Nobuko; Satonaka, Mitsuko; Imagawa, Masayoshi; Naito, Haruna; Kawasaki, Toshisuke
CORPORATE SOURCE: College of Medical Technology, Kyoto University, Kyoto, 606-8507, Japan
SOURCE: Journal of Biochemistry (Tokyo) (1998), 124(6), 1188-1197
CODEN: JOBIAO; ISSN: 0021-924X
PUBLISHER: Japanese Biochemical Society
DOCUMENT TYPE: Journal
LANGUAGE: English

AB Bovine conglutinin is a Ca2+ -dependent serum lectin that is specific for N-acetylglucosamine and a member of the **collectin** (collagen-like

lectin) family. Here we report the identification of the cis-acting elements involved in regulating expression of the conglutinin **gene** . The 5'-flanking region of the conglutinin **gene** was cloned and sequenced by **gene** walking using vector (cassette)-ligation mediated PCR. A genomic fragment encompassing -741 to +50 bp had significant promoter activity when linked to the luciferase reporter **gene** and transfected into the human hepatoma cell line HepG2. Transfection anal. using a series of luciferase vector/5'-stepwise deletion mutants of the promoter constructs indicated that the sequence of 7 base pairs at around -180 bp from the transcription initiation site was necessary for the full expression of the conglutinin **gene**. The site-directed mutagenesis in the AP-1 (Activator **Protein**-1) sequence, immediately down-stream of the pos. controlling cis-element at around -180 bp, resulted in a marked loss of the promoter activity. The novel pos. controlling cis-element and the AP-1 sequence regulated synergistically the expression of the conglutinin **gene**. Gel retardation assay and DNase I footprint anal. demonstrated the presence of the nuclear **proteins** that bind to these two cis-elements.

REFERENCE COUNT: 44 THERE ARE 44 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 34 OF 57 HCAPLUS COPYRIGHT 2002 ACS

ACCESSION NUMBER: 1998:704632 HCAPLUS

DOCUMENT NUMBER: 130:64058

TITLE: Surfactant **protein**-D regulates surfactant phospholipid homeostasis in Vivo

AUTHOR(S): Korfhagen, Thomas R.; Sheftelyevich, Vladimir; Burhans, Michael S.; Bruno, Michael D.; Ross, Gary F.; Wert, Susan E.; Stahlman, Mildred T.; Jobe, Alan H.; Ikegami, Machiko; Whitsett, Jeffrey A.; Fisher, James H.

CORPORATE SOURCE: Division of Pulmonary Biology, Department of Pediatrics, Children's Hospital Research Foundation, Cincinnati, OH, 45229-3039, USA

SOURCE: Journal of Biological Chemistry (1998), 273(43), 28438-28443
CODEN: JBCHA3; ISSN: 0021-9258

PUBLISHER: American Society for Biochemistry and Molecular Biology

DOCUMENT TYPE: Journal

LANGUAGE: English

AB Surfactant **protein** D (SP-D) is a 43-kDa member of the **collectin** family of collagenous lectin domain-contg. **proteins** that is expressed in epithelial cells of the lung. The SP-D **gene** was targeted by homologous recombination in embryonic stem cells that were used to produce SP-D (.+-.) and SP-D (-/-) mice. Both SP-D (-/-) and SP-D (.+-.) mice survived normally in the perinatal and postnatal periods. Whereas no abnormalities were obsd. in SP-D (.+-.) mice, alveolar and tissue phosphatidylcholine pool sizes were markedly increased in SP-D (-/-) mice. Increased nos. of large foamy alveolar macrophages and enlarged alveoli were also obsd. in SP-D (-/-) mice. Phospholipid compn. was unaltered in SP-D (-/-) mice, but surfactant morphol. was abnormal, consisting of dense phospholipid membranous arrays with decreased tubular myelin. The pulmonary lipoidosis in the SP-D (-/-) mice was not assocd. with accumulation of surfactant **proteins** B or C, or their mRNAs, distinguishing the disorder from alveolar proteinosis syndromes. Surfactant **protein** A mRNA was reduced and, SP-A **protein** appeared to be reduced in SP-D (-/-) compared with wild type mice. Targeting of the mouse SP-D **gene** caused

accumulation of surfactant lipid and altered phospholipid structures, demonstrating a previously unsuspected role for SP-D in surfactant lipid homeostasis in vivo.
REFERENCE COUNT: 28 THERE ARE 28 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 35 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 1998:695499 HCAPLUS
DOCUMENT NUMBER: 130:94022
TITLE: Mannose-binding lectin (MBL) in health and disease
AUTHOR(S): Turner, Malcolm W.
CORPORATE SOURCE: Immunobiology Unit, Institute of Child Health, London, UK
SOURCE: Immunobiology (1998), 199(2), 327-339
CODEN: IMMND4; ISSN: 0171-2985
PUBLISHER: Gustav Fischer Verlag
DOCUMENT TYPE: Journal; General Review
LANGUAGE: English
AB A review with 48 refs. Mannose-binding lectin (MBL) is the most intensively studied human **collectin**. It is recognized to be a versatile macro-mol. with many of the functional characteristics of IgM, IgG and Clq. In the presence of calcium the **protein** can bind to a wide spectrum of oligosaccharides through multiple lectin domains. Such binding to the repeating sugar arrays on microbial surfaces may result in direct uptake by one or more **collectin** receptors on phagocyte surface or may trigger the activation of a pro-serine protease complex (MASP 1 and MASP 2) leading to cleavage of C4 and C2 of the classical complement pathway. Although serum levels of MBL are normally rather low (1500 .mu.g/L) there is increasing evidence that the **protein** plays an important role in immune defense, particularly during the phase of primary contact with a microorganism. This is suggested by the obsd. assocn. of an increased incidence of infections in individuals with structural mutations in exon 1 of the MBL **gene**. A cluster of such mutations in codons 52, 54 and 57 lead to secondary structural abnormalities of the collagenous triple helix and a failure to form biol. functional higher order oligomers. The codon 54 mutation has been identified in several Eurasian populations whereas the codon 57 mutation is characteristic of sub-Saharan populations. One intriguing paradox arising from the MBL genotyping studies is the observation that in many populations there are surprisingly high frequencies of either the codon 54 or codon 57 mutation, suggesting that there may be some biol. advantage assocd. with absence of the **protein**. Nevertheless, various groups have reported either low serum levels of MBL or an increased frequency of the structural **gene** mutations in patients with suspected immunodeficiencies, those with frequent unexplained infections and those with systemic lupus erythematosus. There is also evidence that the rate of progression of AIDS in HIV pos. men is faster in those with such mutations. A recently published study of a consecutive series of admissions to a pediatric unit suggests that children presenting with an infectious etiol. are significantly more likely to have a MBL mutation. Moreover, this assocn. was independent of age. Prospective studies are underway to address the questions raised by these findings.
REFERENCE COUNT: 49 THERE ARE 49 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 36 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 1998:649230 HCAPLUS
DOCUMENT NUMBER: 130:12721
TITLE: Altered surfactant homeostasis and alveolar type II

cell morphology in mice lacking surfactant **protein** D
AUTHOR(S): Botas, Carlos; Poulain, Francis; Akiyama, Jennifer; Brown, Cindy; Allen, Lennell; Goerke, Jon; Clements, John; Carlson, Elaine; Gillespie, Anne Marie; Epstein, Charles; Hawgood, Samuel
CORPORATE SOURCE: Cardiovascular Research Institute and Department of Pediatrics, University of California, San Francisco, CA, 94118-1245, USA
SOURCE: Proceedings of the National Academy of Sciences of the United States of America (1998), 95(20), 11869-11874
CODEN: PNASA6; ISSN: 0027-8424
PUBLISHER: National Academy of Sciences
DOCUMENT TYPE: Journal
LANGUAGE: English
AB Surfactant **protein** D (SP-D) is one of two **collectins** found in the pulmonary alveolus. On the basis of homol. with other **collectins**, potential functions for SP-D include roles in innate immunity and surfactant metab. The SP-D **gene** was disrupted in embryonic stem cells by homologous recombination to generate mice deficient in SP-D. Mice heterozygous for the mutant SP-D allele had SP-D concns. that were approx. 50% wild type but no other obvious phenotypic abnormality. Mice totally deficient in SP-D were healthy to 7 mo but had a progressive accumulation of surfactant lipids, SP-A, and SP-B in the alveolar space. By 8 wk the alveolar phospholipid pool was 8-fold higher than wild-type littermates. There was also a 10-fold accumulation of alveolar macrophages in the null mice, and many macrophages were both multinucleated and foamy in appearance. Type II cells in the null mice were hyperplastic and contained giant lamellar bodies. These alterations in surfactant homeostasis were not assocd. with detectable changes in surfactant surface activity, postnatal respiratory function, or survival. The findings in the SP-D-deficient mice suggest a role for SP-D in surfactant homeostasis.
REFERENCE COUNT: 36 THERE ARE 36 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L6 ANSWER 37 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 1998:233761 HCAPLUS
DOCUMENT NUMBER: 129:66588
TITLE: Molecular and biological characterization of rabbit mannan-binding **protein** (MBP)
AUTHOR(S): Kawai, Takao; Suzuki, Yasuhiko; Eda, Souji; Ohtani, Katsuki; Kase, Tetsuo; Sakamoto, Takashi; Uemura, Hidetoshi; Wakamiya, Nobutaka
CORPORATE SOURCE: Department of Food Microbiology, Osaka Prefectural Institute of Public Health, Osaka, 537, Japan
SOURCE: Glycobiology (1998), 8(3), 237-244
CODEN: GLYCE3; ISSN: 0959-6658
PUBLISHER: Oxford University Press
DOCUMENT TYPE: Journal
LANGUAGE: English
AB Mannan-binding **protein** (MBP) is a member of the **collectin** family of **protein**. There are two types of MBP, MBP-A and MBP-C, which were found in rodent (rats and mice), rhesus monkey, and cynomolgus monkey, while chimpanzee and human have only one MBP. It was considered that the loss of one MBP **gene** occurred during hominoid evolution. In this article two rabbit MBP, a liver and serum MBP, were characterized biol. and genetically. Analyses by SDS-PAGE under reduced condition and their amino acid sequences of both MBPs showed

that they have a same mol. wt. of 32 kDa and their amino acid sequences were identical. A serum MBP has a higher ability to activate complement than does a liver MBP; however, a liver MBP inhibits hemagglutination by influenza virus as strongly as a serum MBP does. The cDNA clones encoding the rabbit MBP were isolated from a rabbit cDNA liver library using whole cDNA of mouse MBP-C as a probe. The cDNA carried an insert of 744 bp coding for a **protein** of 247 acid residues with a signal peptide of 22 residues. The deduced amino acid sequence of the cDNA was identical to that of amino acid sequences of the 32-kDa **proteins** detd. here. Northern blot anal. showed that mRNA transcripts of about 0.9 and 3.0 kb were expressed only in the liver. The anal. of the phylogenetic tree of rabbit and bovine MBPs and other **collectins** indicates that the loss of MBP **gene** occurred not only during hominoid evolution but also at some points after the sepn. of birds and mammals.

L6 ANSWER 38 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 1998:25191 HCAPLUS
DOCUMENT NUMBER: 128:73944
TITLE: The role of **collectins** in host defense
AUTHOR(S): Sumiya, Michiko; Summerfield, John A.
CORPORATE SOURCE: Liver Unit, Imperial College School of Medicine at St. Mary's, London, UK
SOURCE: Seminars in Liver Disease (1997), 17(4), 311-318
CODEN: SLDIEE; ISSN: 0272-8087
PUBLISHER: Thieme Medical Publishers, Inc.
DOCUMENT TYPE: Journal; General Review
LANGUAGE: English
AB A review with 98 refs. Mannose-binding **protein** (MBP) belongs to a group of Ca+2-dependent lectins called **collectins** that play a role in first-line host defense. It recognizes specific carbohydrate residues (mannose and N-acetylglucosamine) on the surface of microorganisms and promotes the killing of microbes either by acting directly as an opsonin or by activating the lectin complement pathway. The collagen-like domain of MBP is important for the binding of MBP to the **collectin** receptors expressed on different phagocytes, and for activation of complement. The binding of MBP to bacteria, viruses, and parasites has been demonstrated in vitro. Three major mutations have been found in exon 1 of the MBP **gene**, which encodes the collagenous domain of the **protein**. These mutations cause low levels of serum MBP and have been linked with life-long risk of infection. The homozygotes for these mutations are esp. susceptible to severe infections.

L6 ANSWER 39 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 1997:728261 HCAPLUS
DOCUMENT NUMBER: 128:2689
TITLE: Immunomodulatory functions of surfactant
AUTHOR(S): Wright, Jo Rae
CORPORATE SOURCE: Department of Cell Biology, Duke University School of Medicine, Durham, NC, USA
SOURCE: Physiological Reviews (1997), 77(4), 931-962
CODEN: PHREA7; ISSN: 0031-9333
PUBLISHER: American Physiological Society
DOCUMENT TYPE: Journal; General Review
LANGUAGE: English
AB A review with 298 refs. The possibility that the lipoprotein complex of lung surfactant functions in pulmonary host defense as well as lowering surface tension at the air-liq. interface has been the subject of renewed interest in light of the finding that surfactant **proteins** A and D (SP-A and SP-D) are members of a family of **proteins** known as

collectins. The **collectins**, so named because they have in common an NH2-terminal collagen-like domain and a COOH-terminal lectin (carbohydrate binding) domain, are found in both lung and serum and participate in "innate" immunity, acting before induction of an antibody-mediated response. In vitro, many of the **collectins** stimulate phagocytosis, chemotaxis, and prodn. of reactive oxygen and regulate cytokine release by immune cells. It has been known for several years that surfactant lipids suppress a variety of immune cell functions, most notably lymphocyte proliferation, which, conversely, is augmented by SP-A. Thus surfactant lipids and **proteins** may be counterregulatory, and changes in lipid-to-**protein** ratios may be important in regulating the immune status of the lung. That these ratios change in disease states is clear, but it is not known whether the alterations are a cause or an effect. Important future studies with mice in which the SP-A and SP-D **genes** have been ablated will help clarify the role of surfactant in immune function.

L6 ANSWER 40 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 1997:459846 HCAPLUS
DOCUMENT NUMBER: 127:187150
TITLE: Tetranectin, a trimeric plasminogen-binding C-type lectin
AUTHOR(S): Holtet, Thor Las; Graversen, Jonas Heilskov; Clemmensen, Inge; Thoegersen, Hans Christian; Etzerodt, Michael
CORPORATE SOURCE: Laboratory of Gene Expression, Department of Molecular and Structural Biology, University of Aarhus, Aarhus, DK-8000, Den.
SOURCE: Protein Science (1997), 6(7), 1511-1515
CODEN: PRCIEI; ISSN: 0961-8368
PUBLISHER: Cambridge University Press
DOCUMENT TYPE: Journal
LANGUAGE: English
AB Tetranectin, a plasminogen-binding **protein** belonging to the family of C-type lectins, was expressed in Escherichia coli and converted to its native form by in vitro refolding and proteolytic processing. Recombinant tetranectin, as well as natural tetranectin from human blood plasma, was shown by chem. crosslinking anal. and SDS-PAGE to be a homotrimer in soln. as are other known members of the **collectin** family of C-type lectins. Biochem. evidence is presented showing that an N-terminal domain encoded within exons 1 and 2 of the tetranectin **gene** is necessary and sufficient to govern subunit trimerization.

L6 ANSWER 41 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 1997:393264 HCAPLUS
DOCUMENT NUMBER: 127:79841
TITLE: **Collectins**
AUTHOR(S): Sastry, Kedarnath N.; Ezekowitz, R. Alan B.
CORPORATE SOURCE: Boston University School of Medicine, Boston, MA, USA
SOURCE: Collectins and Innate Immunity (1996), 1-7.
Editor(s): Ezekowitz, R. Alan B.; Sastry, Kedarnath N.; Reid, Kenneth B. M. Landes: Austin, Tex.
CODEN: 64OBA7
DOCUMENT TYPE: Conference; General Review
LANGUAGE: English
AB A review with 18 refs. discussing classification and structure of **collectins**, biol. activities, and evolution and organization of **collectin genes**.

L6 ANSWER 42 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 1996:626218 HCAPLUS
DOCUMENT NUMBER: 125:273548
TITLE: Biosynthesis of human ficolin, an Escherichia coli-binding **protein**, by monocytes: comparison with the synthesis of two macrophage-specific **proteins**, C1q and the mannose receptor
AUTHOR(S): Lu, J.; Le, Y.; Kon, L.; Chan, J.; Lee, S. H.
CORPORATE SOURCE: Dep. Biochem., Faculty Medicine, National Univ. Singapore, Singapore, Singapore
SOURCE: Immunology (1996), 89(2), 289-294
CODEN: IMMUAM; ISSN: 0019-2805
DOCUMENT TYPE: Journal
LANGUAGE: English
AB Ficolin is characterized by the presence of both collagen-like and fibrinogen-like sequences, and potentially has a similar overall structure as the complement **protein** C1q and the **collectins**. Previous studies have reported the presence of human ficolin mRNA predominantly in peripheral blood leukocytes. In the present study, the cellular origin of human ficolin was investigated in further detail. Preliminary studies using reverse transcriptase-polymerase chain reaction (RT-PCR) showed that ficolin mRNA was synthesized by U937 cells, a human monocyte cell line. This finding suggested that blood monocytes also normally synthesize human ficolin. Peripheral blood monocytes from adult human donors were harvested at serial time-points (0-20 h) after adhesion to tissue culture plates, and total **RNA** was isolated and assayed for ficolin mRNA by RT-PCR. Ficolin mRNA was highly expressed in monocytes throughout the first 20 h of adhesion. In contrast, C1q and mannose receptor mRNA were not detectable during the first 8 h of adhesion, but were highly expressed by 20 h. Cells were harvested at longer time intervals (1, 2, 4, 6 and 8 days) to det. whether ficolin expression was temporally regulated at later stages of monocyte differentiation. Ficolin mRNA levels decreased sharply from day 1 to day 6. In contrast, the levels of both C1q and mannose receptor mRNA showed no changing trend. These results are consistent with the absence of ficolin expression in many macrophage-rich tissues previously reported. The origin of ficolin from monocytes, together with its structural similarity to C1q and the **collectins**, raises the possibility that ficolin may be another plasma **protein** capable of binding to surface structures of micro-organisms. ,. Escherichia coli was therefore incubated with human serum, and bound **proteins**, after elution with sugars, were analyzed by Western blotting using an antiserum raised against a synthetic ficolin peptide. The antiserum identified a polypeptide of approx. 42,000 MW, which is similar in size to that of ficolin as predicted from its cDNA-derived sequence.

L6 ANSWER 43 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 1996:568047 HCAPLUS
DOCUMENT NUMBER: 125:266944
TITLE: Characterization of two mannose-binding **protein** cDNAs from rhesus monkey Macaca mulatta: structure and evolutionary implications
AUTHOR(S): Mogues, Tirsit; Ota, Tatsuya; Tauber, Alfred I.; Sastry, Kedarnath N.
CORPORATE SOURCE: Departments of Pathology and Medicine, Boston University, Boston, MA, 02118, USA
SOURCE: Glycobiology (1996), 6(5), 543-550
CODEN: GLYCE3; ISSN: 0959-6658

DOCUMENT TYPE: Journal
LANGUAGE: English
AB Mannose-binding **proteins** (MBPs), members of the **collectin** family, have been implicated as lectin opsonins for various viruses and bacteria. Two distinct but related MBPs, MBP-A and MBP-C, with .apprx.55% identity at the amino acid level, have been previously characterized from rodents. In humans, however, only one form of MBP has been characterized. In this paper we report studies elucidating the evolution of primate MBPs. ELISA and Western blot analyses indicated that rhesus and cynomolgus monkeys have two forms of MBP in their sera, whereas chimpanzees have only one form, similar to humans. Two distinct MBP cDNA clones were isolated and characterized from a rhesus monkey liver cDNA library. Rhesus MBP-Ais closely related to the mouse and rat MBP-A, showing 77% and 75% identity at the amino acid level, resp. Rhesus MBP-A also has three cysteines at the N-terminus, similar to mouse and rat MBP-A and human MBP. Rhesus MBP-C shares 90% identity with the human MBP at the amino acid level and has three cysteines at the N-terminus, in contrast to two cysteine residues found in rodent MBP-C. A stretch of nine amino acids close to the N-terminus, absent in both mouse and rat MBP-A, but present in rodent MBP-C, chicken and human MBPs, is also found in the rhesus MBP-A. The phylogenetic anal. of rhesus and other mammalian MBPs, coupled with the serol. data suggest that at least two distinct MBP **genes** existed prior to mammalian radiation and the hominoid ancestor apparently lost one of these **genes** or failed to express it.

L6 ANSWER 44 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 1996:438828 HCAPLUS
DOCUMENT NUMBER: 125:111012
TITLE: Localization and development expression of surfactant **proteins** D and A in the respiratory tract of the mouse
AUTHOR(S): Wong, Carlene J.; Akiyama, Jennifer; Allen, Lennell; Hawgood, Samuel
CORPORATE SOURCE: San Francisco School Medicine, University California, San Francisco, CA, 94143-0130, USA
SOURCE: Pediatr. Res. (1996), 39(6), 930-937
CODEN: PEREBL; ISSN: 0031-3998
DOCUMENT TYPE: Journal
LANGUAGE: English
AB Surfactant **protein** D (SP-D) is synthesized and secreted by pulmonary epithelial cells. Like surfactant **protein** A (SP-A), SP-D is a collagen-like glycoprotein belonging to the "**collectin** " class of C-type lectins that may play an important role in pulmonary host defense. To begin studies on SP-D **gene** regulation and function using the mouse as an animal model, the authors identified the cellular sites of SP-D **gene** expression in adult mouse lung and trachea and characterized the developmental expression of SP-D mRNA in murine fetal and newborn lungs. The authors compared these findings with similar studies for murine SP-A, which has an established role in surfactant function and metab. and a probable role in pulmonary host defense. SP-D mRNA and **protein** were readily detected by in situ hybridization and immunocytochem. in alveolar type II and nonciliated bronchiolar epithelial cells of the lung, as well as in cells of the tracheal epithelium and tracheal submucosal glands of the adult mouse. Although SP-A mRNA and **protein** were also localized to alveolar and nonciliated bronchiolar epithelial cells of the murine lung, there was no detectable labeling for either SP-A mRNA or **protein** in the murine trachea. Expression of murine SP-D mRNA was first detected by

Northern blot anal. on d 16 of gestation in timed-pregnant mice, with an av. gestational period of 17 d, and this increased dramatically before birth and during the immediate postnatal period. The development expression of murine SP-A mRNA paralleled that of SP-D except that there was a small decrease in mRNA content on postnatal d 5. These studies provide the first description of the cellular distribution and developmental expression of SP-D in mouse lung, which will be important for interpreting future studies of SP-D **gene** expression in transgenic animal models. In addn., these studies provide the first documentation that, unlike SP-A, SP-D is synthesized not only in the lung but also in submucosal glands of the trachea.

L6 ANSWER 45 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 1996:257988 HCAPLUS
DOCUMENT NUMBER: 124:286887
TITLE: Mutations in the human mannose-binding **protein** gene: Frequencies in several population groups
AUTHOR(S): Lipscombe, R. J.; Beatty, D. W.; Ganczakowski, M.; Goddard, E. A.; Jenkins, T.; Lau, Y. L.; Spurdle, A. B.; Sumiya, M.; Summerfield, J. A.; Turner, M. W.
CORPORATE SOURCE: Molecular Immunology Unit, Institute Child Health, London, WC1N 1EH, UK
SOURCE: Eur. J. Hum. Genet. (1996), 4(1), 13-19
CODEN: EJHGEU; ISSN: 1018-4813
DOCUMENT TYPE: Journal
LANGUAGE: English
AB Mannose-binding **protein** (MBP; mannan-binding **protein**, mannan-binding lectin) is a member of the **collectin** family of **proteins** and is thought to be important in innate immunity. The authors have previously shown high frequencies of two distinct mutations in codon 54 and codon 57 of exon 1 of the MBP **gene** in non-African and African populations, resp. These result in low levels of the **protein** and an opsonic deficiency but the frequencies also suggest some selective advantage for low MBP levels. A third mutation in codon 52 occurs at a much lower frequency. The authors have now extended their earlier studies to other populations. In the south-west Pacific (Papua New Guinea and Vanuatu) neither the codon 52 nor the codon 57 mutation was detected and the codon 54 mutation was significantly less common (**gene** frequencies of 0.07 and 0.01, resp.) than in other non-African populations (**gene** frequencies 0.11-0.16). This could be explained by relatively recent admixt. The ancestral Melanesian population probably diverged some 50,000-60,000 yr ago and the authors' data suggest that the codon 54 mutation may have occurred after that event but before the divergence of European-Asian groups (40,000 yr ago). Two further sub-Saharan populations were also studied: a group of Xhosa from South Africa were similar to Gambians, with a high **gene** frequency for the codon 57 mutation (0.27) and no evidence of the codon 52 or 54 mutations. In contrast, San Bushmen from Namibia had low frequencies of both the codon 57 mutation (0.07) and the codon 54 mutation (0.03). Again the codon 52 mutation was not found. This pattern is unique amongst sub-Saharan populations studied to date and suggests that this population may have been subjected to different selective pressures.

L6 ANSWER 46 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 1996:64971 HCAPLUS
DOCUMENT NUMBER: 124:108951
TITLE: Trimerizing polypeptides, their manufacture using cloning vectors and use
INVENTOR(S): Hoppe, Hans-Juergen; Reid, Kenneth Bannerman Milne

PATENT ASSIGNEE(S): Medical Research Council, UK
SOURCE: PCT Int. Appl., 81 pp.
CODEN: PIXXD2
DOCUMENT TYPE: Patent
LANGUAGE: English
FAMILY ACC. NUM. COUNT: 1
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 9531540	A1	19951123	WO 1995-GB1104	19950516
W: AM, AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LT, LU, LV, MD, MG, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TT				
RW: KE, MW, SD, SZ, UG, AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG				
CA 2190264	AA	19951123	CA 1995-2190264	19950516
AU 9524519	A1	19951205	AU 1995-24519	19950516
EP 757720	A1	19970212	EP 1995-918689	19950516
R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LI, LU, MC, NL, PT, SE				
JP 10500298	T2	19980113	JP 1995-529462	19950516
US 6190886	B1	20010220	US 1997-737629	19970110

PRIORITY APPLN. INFO.: GB 1994-9768 A 19940516
WO 1995-GB1104 W 19950516

AB Polypeptides comprising a **collectin** neck region, or variant or deriv. thereof or amino acid sequence having the same or a similar amino acid pattern and/or hydrophobicity profile, are able to trimerize. Such polypeptides may comprise addnl. amino acids which may include heterologous amino acids, for example forming a **protein** domain or derived from an Ig or comprising an amino acid which may be derivatized for attachment of a non-peptide moiety such as oligosaccharide, and may form homotrimers or heterotrimers. Heterotrimerization may be promoted by gentle heating, e.g. to about 50.degree.C, then cooling to room temp. One use for the polypeptides is in seeding collagen formation. **Nucleic acid** encoding the polypeptides and methods fo their prodn. are provided.

L6 ANSWER 47 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 1995:986217 HCAPLUS
DOCUMENT NUMBER: 124:113133
TITLE: Structure-function relationships in the calcium-dependent animal lectins
AUTHOR(S): Uemura, Kazuhide; Kawasaki, Nobuko; Kawasaki, Toshisuke
CORPORATE SOURCE: Fac. Pharm. Sci., Kyoto Univ., Kyoto, 606-01, Japan
SOURCE: Jikken Igaku (1995), 13(18), 2156-61
CODEN: JIIGEF; ISSN: 0288-5514
DOCUMENT TYPE: Journal; General Review
LANGUAGE: Japanese

AB A review, with 25 refs., on the point mutation of serum mannan-binding **protein gene** and opsonin insufficiency, the structure-function relations of **collectin** based on the **gene** structure of bovine serum conglutinin, and C-type lectin-mediated intracellular signal transduction.

L6 ANSWER 48 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 1995:983903 HCAPLUS

DOCUMENT NUMBER: 124:53627
TITLE: Mouse surfactant **protein**-D. cDNA cloning, characterization, and gene localization to chromosome 14
AUTHOR(S): Motwani, Monica; White, Robert A.; Guo, Ning; Dowler, Lisa L.; Tauber, Alfred I.; Sastry, Kednarth, N.
CORPORATE SOURCE: Sch. Med., Boston Univ., Boston, MA, 02118, USA
SOURCE: J. Immunol. (1995), 155(12), 5671-7
CODEN: JOIMA3; ISSN: 0022-1767
DOCUMENT TYPE: Journal
LANGUAGE: English

AB Surfactant **protein**-D (SP-D) is a **collectin** found assocd. with surfactant in the lung. SP-D has also been functionally characterized as an opsonin for diverse microorganisms and chemoattractant for phagocytic cells. To det. the structure of mouse SP-D, the authors isolated and characterized clones from a B6/CBAF1J strain lung cDNA library using a PCR-derived genomic probe. The deduced sequence predicts a 19-amino acid signal sequence, a 25-amino acid long NH2 terminus with two cysteines, followed by an uninterrupted collagen domain with 59 Gly-X-Y repeats. Next, a short "neck" domain of 28 amino acids, with a potential to form trimeric .alpha.-helical coiled coil is found ending in a COOH-terminal 125-amino acid carbohydrate recognition domain. The mature mouse SP-D **protein** of 355 amino acids shows strong homol. to rat (92% identity), human (76%), and bovine (72%) SP-D amino acid sequences. Northern blot and RT-PCR anal. revealed that the mouse SP-D **gene** is expressed predominantly in lung and, surprisingly, also in heart, stomach, and kidney but not in brain. In contrast, mouse surfactant **protein**-A (SP-A) mRNA expression was restricted to lung. Human lung and stomach, but not heart or liver were found to express SP-D mRNA, as detd. by PCR. The mouse SP-D **gene** (Sftp4) has been localized to chromosome 14 (to a region syntenic to human chromosome 10), closely linked to the **genes** for other collagenous lectins, mannose-binding **protein**-A (Mbl1), and Sp-A (Sftp1).

L6 ANSWER 49 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 1995:768605 HCAPLUS
DOCUMENT NUMBER: 123:167591
TITLE: Distinct physicochemical characteristics of human mannose binding **protein** expressed by individuals of differing genotype
AUTHOR(S): Lipscombe, R. J.; Sumiya, M.; Summerfield, J. A.; Turner, M. W.
CORPORATE SOURCE: Mol. Immunol. Unit, Imperial Coll. Sci. Technol., London, UK
SOURCE: Immunology (1995), 85(4), 660-7
CODEN: IMMUAM; ISSN: 0019-2805
DOCUMENT TYPE: Journal
LANGUAGE: English

AB Mannose binding **protein** (MBP) is a serum **collectin** (collagenous lectin) believed to be of importance in innate immunity. Three point mutations, in codons 52, 54 and 57 of exon 1 of the human MBP **gene**, have been predicted to affect the tertiary structure of the collagenous region of the **protein**, and are known to be assocd. with low serum concns. of MBP. However, other groups working with recombinant mutant **proteins** have claimed that the **proteins** are expressed and assembled normally. The aim of the present investigation was to characterize the effects of these mutations on the physicochem. nature of MBP that is present in the circulation in

vivo, and for this we used polyacrylamide gel electrophoresis, gel filtration and sucrose d. gradient centrifugation, followed by immunoblotting and enhanced chemiluminescence. The circulating wild-type MBP appeared to comprise a mixt. of polymers formed from two to eight subunits (each based on three identical 32,000 MW polypeptide chains) of apparent mol. wts. 200,000-700,000, with dimers and trimers constituting the predominant forms. Individuals homozygous for the codon 54 or 57 mutation had dramatically reduced concns. of serum MBP, mainly comprising material of an apparent mol. wt. of 120,000-130,000. Heterozygous individuals showed characteristics of both phenotypes. In contrast to the results obtained with artificial expression systems, our data suggest that individuals homozygous for the MBP mutations have very little circulating **protein** and that this comprises mainly low mol. wt. material.

L6 ANSWER 50 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 1995:564723 HCAPLUS
DOCUMENT NUMBER: 123:103916
TITLE: Characterization of murine mannose-binding **protein genes** Mbl1 and Mbl2 reveals features common to other **collectin genes**
AUTHOR(S): Sastry, R.; Wang, J. -S.; Brown, D. C.; Ezekowitz, R. A. B.; Tauber, A. I.; Sastry, K. N.
CORPORATE SOURCE: Department Pathology, Boston University School Medicine, Boston, MA, 02118, USA
SOURCE: Mamm. Genome (1995), 6(2), 103-10
CODEN: MAMGEC; ISSN: 0938-8990
DOCUMENT TYPE: Journal
LANGUAGE: English

AB Mannose-binding **protein** (MBP) is a member of a family of collagenous lectins (**collectins**), which are believed to play an important role in first-line host defense. In this study, the two **genes** encoding MBP in mice-Mbl1 and Mbl2-have been isolated and their exon-intron structure studied to understand their evolutionary relationship to the single human (MBL) and the two rat MBP **genes**. Mouse Mbl1 and Mbl2 have five and six exons, resp. The structure of the mouse Mbl **genes** is similar to that of the rat and human MBP **genes** and shows homol. to the other **collectin genes**, with the entire carbohydrate recognition domain being encoded in a single exon and all introns being in phase 1. The MBP encoded by mouse Mbl1 with three cysteines in the first coding exon, like the rat Mbl1 and human MBL, is capable of a higher degree of multimerization and has apparent ability to fix complement in the absence of antibody or C1q. However, the structural features of other exons, i.e., the larger size of collagen domain region in the first coding exon (64 bp in Mbl2 vs 46 bp in Mbl1) and the smaller size of the exon encoding the trimerization domain (69 bp in Mbl2 vs 75 bp in Mbl1) reveal that the single human MBL **gene** is closely related to rodent Mbl2 rather than rodent Mbl1. The findings in this study suggest that in contrast to the evolution of another **collectin gene**-bovine surfactant **protein**-D-which duplicated in bovidae after divergence from humans, MBP **gene** most likely duplicated prior to human-rodent divergence, and that the human homolog to Mbl1 was perhaps lost during evolution.

IT **142978-84-5, Protein** RaRF (mouse clone a10 subunit P28a precursor reduced) **142978-86-7, Protein** RaRF (mouse clone b60 subunit P28b precursor reduced)
RL: BOC (Biological occurrence); PRP (Properties); BIOL (Biological study); OCCU (Occurrence)

(amino acid sequence; characterization of murine mannose-binding **protein genes** Mbl1 and Mbl2 reveals features common to other **collectin genes**)

L6 ANSWER 51 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 1995:21509 HCAPLUS
DOCUMENT NUMBER: 122:124552
TITLE: Bovine conglutinin gene exon structure reveals its evolutionary relationship to surfactant **protein**-D
AUTHOR(S): Liou, Louis S.; Sastry, Rajeswari; Hartshorn, Kevan L.; Lee, Young M.; Okarma, Thomas B.; Tauber, Alfred I.; Sastry, Kedarnath N.
CORPORATE SOURCE: Sch. Med., Boston Univ., Boston, MA, 02118, USA
SOURCE: J. Immunol. (1994), 153(1), 173-80
CODEN: JOIMA3; ISSN: 0022-1767
DOCUMENT TYPE: Journal
LANGUAGE: English
AB Bovine conglutinin (BC), a member of the mammalian C-type **collectin** subfamily, is a serum **protein** synthesized in liver that is believed to play a role in natural host defense. Previously, the authors have characterized a full length BC cDNA and the authors now describe the partial characterization of a genomic clone that encodes for the BC **gene** (CGN1). BC is encoded by nine exons spanning >11 kb and has been localized previously to band 18 of bovine (Bos taurus) chromosome 28. Genomic sequencing demonstrated that the signal peptide/amino-terminal domain, the carbohydrate recognition domain, and the linking peptide, a domain between the collagenous region and the carbohydrate recognition domain, are each encoded by a single exon. The collagenous domain is split into five exons, with the 5' most region being located within the exon that also encodes the signal peptide/amino terminus. The remaining four collagenous domain exons are tandemly arranged with lengths of 117, 108, 108, and 117 bp, resp. Overall, the BC genomic organization is very similar to that of the human surfactant **protein**-D **gene**, SFTP4. On the basis of identical collagen domain structures, the authors suggest that conglutinin and bovine surfactant **protein**-D evolved from a **gene** duplication event occurring in Bovidae after divergence from other mammals.

L6 ANSWER 52 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 1994:602561 HCAPLUS
DOCUMENT NUMBER: 121:202561
TITLE: **Collectins**, soluble **proteins** containing collagenous regions and lectin domains, and their roles in innate immunity
AUTHOR(S): Hoppe, Hans-Juergen; Reid, Kenneth B. M.
CORPORATE SOURCE: MRC Immunochemistry Unit, Department Biochemistry, University of Oxford, Oxford, OX1 3QU, UK
SOURCE: Protein Sci. (1994), 3(8), 1143-58
CODEN: PRCIEI; ISSN: 0961-8368
DOCUMENT TYPE: Journal; General Review
LANGUAGE: English
AB A review with 137 refs. The **collectins** are a group of mammalian lectins contg. collagen-like regions. They include mannan binding **protein**, bovine conglutinin, lung surfactant **protein** A, lung surfactant **protein** D, and a newly discovered bovine **protein** named **collectin**-43. These **proteins** share a very similar modular domain compn. and overall 3-dimensional

structure. They also appear to play similar biol. roles in the preimmune defense against microorganisms in both serum and lung surfactant. The close evolutionary relationship between the **collectins** is further emphasized by a common pattern of exons in their genomic structures and the presence of a **gene** cluster on chromosome 10 in humans that contains the **genes** known for the human **collectins**. Studies on the structure/function relationships within the **collectins** could provide insight into the properties of a growing no. of **proteins** also contg. collagenous regions such as C1q, the hibernation **protein**, the .alpha.- and .beta.-ficolins, as well as the membrane acetylcholinesterase and the macrophage scavenger receptor.

L6 ANSWER 53 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 1994:452613 HCAPLUS
DOCUMENT NUMBER: 121:52613
TITLE: Primary structure of bovine **collectin**-43 (CL-43). Comparison with conglutinin and lung surfactant **protein**-D
AUTHOR(S): Lim, Boon Leong; Willis, Anthony C.; Reid, Kenneth B. M.; Lu, Jinhua; Laursen, Steen B.; Jensenius, Jens Christian; Holmskov, Uffe
CORPORATE SOURCE: Dep. Biochem., Univ. Oxford, Oxford, OX1 3QU, UK
SOURCE: J. Biol. Chem. (1994), 269(16), 11820-4
CODEN: JBCHA3; ISSN: 0021-9258
DOCUMENT TYPE: Journal
LANGUAGE: English
AB **Collectin**-43 (CL-43) is a bovine serum **protein** that is composed of subunits of three identical chains, each of which contains a collagen region and a C-type carbohydrate recognition domain; thus, CL-43 belongs to the **collectins** (group III of the C-type lectins). The authors have derived the complete primary sequence of CL-43 using partial **protein** sequencing, cDNA cloning, and reverse transcription-polymerase chain reaction techniques. The primary sequence of CL-43 shows that it contains an N-terminal region of 28 residues, followed by a collagenous domain of 38 repeats of Gly-Xaa-Yaa and then a C-terminal section of 159 residues, contg. a short "neck" region and the carbohydrate recognition domain with the conserved residues found in all C-type lectins. The amino acid sequence of CL-43 showed 74% identity to bovine conglutinin and 70% identity to bovine lung surfactant **protein** D (SP-D), but the collagen region is considerably shorter than the 57 Gly-Xaa-Yaa triplets found in the conglutinin and SP-D. Northern blot anal. showed that CL-43 was only synthesized in bovine liver, with no detectable signal in a variety of other bovine tissues, including lung. No cross-hydridizing signals were detected in mRNA from sheep, human, rat, or mouse liver. Since CL-43 and conglutinin have only been detected in members of bovidae, it is probable that an ancestral **gene** of these two **proteins** was first derived from a SP-D like **gene**, and that this ancestral **gene** duplicated during evolution.
IT **156132-72-8, Collectin**-43 (cattle liver)
RL: PRP (Properties)
(amino acid sequence of)
IT **153058-87-8**
RL: PRP (Properties)
(nucleotide sequence of)

L6 ANSWER 54 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 1994:267512 HCAPLUS

DOCUMENT NUMBER: 120:267512
TITLE: The genomics of soluble **proteins** with collagenous domains: C1q, MBL, SP-A, SP-D, conglutinin, and CL-43
AUTHOR(S): Kolble, K.; Reid, K. B. M.
CORPORATE SOURCE: Dep. Biochem., Univ. Oxford, Oxford, OX1 3QU, UK
SOURCE: Behring Inst. Mitt. (1993), 93(Structure-Function-Relationship of C1q and Collectins C1-Esterases: C1r, C1s and C1-Inhibitor in Health and Disease), 81-6
CODEN: BHIMA2; ISSN: 0301-0457
DOCUMENT TYPE: Journal; General Review
LANGUAGE: English

AB A review and discussion with 36 refs. The **gene** cluster encoding the A, B and C chains of human complement C1q has been localized to 1p34.1-1p36.3, on the short arm of chromosome 1. The C1q mol., although it is not a lectin, shows certain structural and functional similarities to a group of mammalian C-type lectins which contain collagen-like regions. These lectins include the serum **proteins** conglutinin, mannose-binding lectin (MBL) and **Collectin**-43 (CL-43) and the lung surfactant **proteins** A and D (SP-A and SP-D). The **genes** for MBL, SP-A and SP-D have been mapped to human chromosome 10, with at least two expressed SP-A **genes** (SP-AI and SP-AII) forming a cluster with an SP-A pseudogene. Somatic cell hybrid mapping places the human SP-A and SP-D **genes** at 10q22-q23 while MBL is localized at 10q21. Conglutinin and CL-43 have so far only been characterized in the bovine system but if there are human analogs of these **proteins** it seems likely that they will also map to the long arm of chromosome 10.

L6 ANSWER 55 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 1994:262640 HCAPLUS
DOCUMENT NUMBER: 120:262640
TITLE: Cloning of a pH-sensitive K+ channel possessing two transmembrane segments
AUTHOR(S): Suzuki, Makoto; Takahashi, Keiko; Ikeda, Masato; Hayakawa, Hiroshi; Ogawa, Aiichirou; Kawaguchi, Yoshindo; Sakai, Osamu
CORPORATE SOURCE: Dep. Pharmacol., Jichi Med. Sch., Kawachi, 329-04, Japan
SOURCE: Nature (London) (1994), 367(6464), 642-5
CODEN: NATUAS; ISSN: 0028-0836
DOCUMENT TYPE: Journal
LANGUAGE: English

AB The mammalian renal collecting ducts are responsible for secreting potassium ions into the urine and are a major regulatory site for potassium homeostasis, in which a voltage-independent pH-sensitive K+ channel in the apical membrane plays a central role. Here the authors describe a cDNA encoding a novel K+ channel from rabbit renal cortical collecting tubule cells (RACTK1). RACTK1 has the functional characteristics of the apical K+-permeable channel and consists of 284 amino acids, putatively with two transmembrane segments. The sequence of RACTK1, however, shows no homol. to known voltage-dependent or -independent K+ channels, and has a different K+-driving path and regulatory sites. The study of this **protein** should provide insight into K+ homeostasis and diseases of K+ metab.

IT **154837-43-1**, RACTK1 potassium channel (rabbit kidney cortical collecting tubule cell)
RL: PRP (Properties)
(amino acid sequence and transport properties of, transport pH

sensitivity in relation to)
IT **154837-42-0, DNA** (rabbit kidney cortical collecting tubule cell pH-sensitive RACTK1 potassium channel cDNA plus flanks)
RL: PRP (Properties); BIOL (Biological study)
(nucleotide sequence and expression of)

L6 ANSWER 56 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 1994:237365 HCAPLUS
DOCUMENT NUMBER: 120:237365
TITLE: Expression, functional analysis, and in situ hybridization of a cloned rat kidney collecting duct water channel
AUTHOR(S): Ma, Tonghui; Hasegawa, Hajime; Skach, William R.; Frigeri, Antonio; Verkman, A. S.
CORPORATE SOURCE: Cardiovasc. Res. Inst., Univ. California, San Francisco, CA, 94143, USA
SOURCE: Am. J. Physiol. (1994), 266(1, Pt. 1), C189-C197
CODEN: AJPHAP; ISSN: 0002-9513
DOCUMENT TYPE: Journal
LANGUAGE: English
AB The cloning and expression of an apical membrane water channel from rat kidney collecting duct (WCH-CD) homologous to a 28-kDa integral membrane **protein** (CHIP28) was reported recently (K. Fushimi, S. Uchida, Y. Hara, Y. Hirata, F. Marumo, and S. Sasaki. Nature Lond. 361: 549-552, 1993). The authors obtained an .apprx.1.8-kilobase clone from a rat kidney .lambda.gt10 cDNA library by a polymerase chain reaction cloning method, whereas the coding sequence (814 bp, predicted **protein** size of 29 kDa) was identical to that reported. The authors identified an in-frame ATG codon at base pair -123 predicting a **protein** size of 33 kDa, contrary to the predicted **protein** size of 29 kDa. Northern blots probed by cDNAs corresponding to the WCH-CD coding sequence (base pairs +1 to +814) or 5'-untranslated sequence (-403 to -16) reveal a single band at 1.9 kilobases in kidney medulla greater and not in other tissues. MRNA expression was increased strongly by dehydration. Translation and oocyte expression studies were performed to identify the translation start site. The short (base pairs +1 to +814) and long (base pairs -123 to +814) cDNAs were sub-cloned in vector pSP64 contg. the 5'-untranslated Xenopus globin sequence upstream to the ATGs; a 30-base pair c-myc sequence was engineered at the C-terminal for antibody recognition. Water permeability in Xenopus oocytes injected with 50 ng of transcribed cRNA was (in cm/s .times. 10-3) 20 .+-. 3 (short clone), 1.3 .+-. 0.2 (long clone), 11 .+-. 3 (short clone with no globin sequence), 0.7 .+-. 0.1 (water-injected control), and 20 .+-. 4 (CHIP28k); the increased water permeability in oocytes expressing the short clone was inhibited by 75% by 0.3 mM HgCl2 but not affected by adenosine 3',5'-cyclic monophosphate agonists. Cell-free translation of the short clone gave a band at 29 kDa that became glycosylated (32 kDa) in the presence of pancreatic microsomes; translation of the long clone was much less efficient. Translation in oocytes followed by anti-c-myc immunopptn. and [35S]methionine autoradiog. gave major bands at 29 and 32 kDa for the short clone. In situ hybridization of rat kidney using a 35S-labeled 187-base cRNA anti-sense probe (base pairs +343 to +529) showed localization of mRNA encoding WCH-CD only to medullary and cortical collecting ducts. These studies indicate that WCH-CD is a collecting duct water channel and provide translation and expression data indicating that the second ATG codon is the major translation initiation site.
IT **154432-61-8**, Kidney collecting duct water channel (Rat **gene** WCH-CD)
RL: PRP (Properties)

(amino acid sequence and in situ hybridization of and functional anal. of)

L6 ANSWER 57 OF 57 HCAPLUS COPYRIGHT 2002 ACS
ACCESSION NUMBER: 1993:468700 HCAPLUS
DOCUMENT NUMBER: 119:68700
TITLE: The cDNA cloning of conglutinin and identification of liver as a primary site of synthesis of conglutinin in members of the Bovidae
AUTHOR(S): Lu, Jinhua; Laursen, Steen B.; Thiel, Steffen; Jensenius, Jens C.; Reid, Kenneth B. M.
CORPORATE SOURCE: Dep. Biochem., Univ. Oxford, Oxford, OX1 3QU, UK
SOURCE: Biochem. J. (1993), 292(1), 157-62
CODEN: BIJOAK; ISSN: 0306-3275
DOCUMENT TYPE: Journal
LANGUAGE: English
AB Bovine conglutinin is a collagen-like, C-type, plasma lectin which belongs to the group of **proteins** called **collectins**. Two inosine-contg. oligonucleotides were synthesized, based on the published **protein** sequence for bovine conglutinin, and PCR on target **DNA** from a bovine liver .lambda.gt 11 cDNA library yielded a product of the expected size of 210 bp. Screening of the library with this cDNA fragment identified a single pos. clone, with an insert of 0.9 kb, coding for bovine conglutinin from residue 70 to the C-terminus. The 5' cDNA sequence, encompassing 150 bp of the 5' non-translated sequence plus the sequence encoding the leader peptide and the N-terminal residues 1-70, was completed by the use of PCR techniques. The cDNA sequence of bovine conglutinin showed 86% identity with that of bovine lung surfactant **protein** D (SP-D), and the derived amino acid sequence of bovine conglutinin showed 78% identity with that of bovine SP-D, which included complete identity of the leader-peptide sequences. The amino acid sequence derived from the cDNA sequence differs from the published **protein** sequence at 4 positions. Northern-blot anal. on total **RNA**, purified from various tissues from cattle, sheep, humans, rats, and mice, showed that a strong signal of .apprx.1.8 kb is present in bovine liver **RNA**. A weak signal of similar size was also obsd. in sheep liver, but not in human, rat, and mouse livers. A weak signal, also of 1.8 kb, is present in the lung **RNAs** of all the species tested. The signals from the lung tissues are likely to be due to the cross-hybridization of the bovine conglutinin cDNA to the SP-D mRNAs of the resp. species. The finding of significant signals in only the bovine and sheep liver **RNA** samples is indicative that serum conglutinin may be present in significant amts. only in members of the Bovidae (the family encompassing cattle, antelopes, sheep, and goats) and closely related species.

```
=> d stat que
L1          77 SEA FILE=REGISTRY ABB=ON  PLU=ON  COLLECTIN/BI
L2         260 SEA FILE=HCAPLUS ABB=ON  PLU=ON  L1 OR COLLECTIN
L4         223 SEA FILE=HCAPLUS ABB=ON  PLU=ON  L2 AND PROTEIN
L5          62 SEA FILE=HCAPLUS ABB=ON  PLU=ON  L2(L)(GENE OR DNA OR NUCLEIC(W
               )ACID OR RNA)
L6          57 SEA FILE=HCAPLUS ABB=ON  PLU=ON  L5 AND L4
L7          38 SEA FILE=HCAPLUS ABB=ON  PLU=ON  (L2(L)(?BACTERI? OR ?STAT? OR
               ?CIDAL OR ?CIDE?)) NOT L6
L8           1 SEA FILE=HCAPLUS ABB=ON  PLU=ON  L5 AND L7
```

```
=>
=>

=> d ibib abs hitrn 18 1

L8   ANSWER 1 OF 1  HCAPLUS  COPYRIGHT 2002 ACS
ACCESSION NUMBER:        2002:431258  HCAPLUS
TITLE:                   Macrophage defences against respiratory tract
                         infections
AUTHOR(S):               Gordon, S. B.; Read, R. C.
CORPORATE SOURCE:        Wellcome Trust Research Laboratories, Queen Elizabeth
                         Central Hospital, Universities of Malawi and Liverpool
                         (UK), Blantyre, Malawi
SOURCE:                  British Medical Bulletin (2002), 61, 45-61
                         CODEN: BMBUAQ; ISSN: 0007-1420
PUBLISHER:               Oxford University Press
DOCUMENT TYPE:           Journal
LANGUAGE:                English
AB   Pulmonary macrophages with a key role in defense against respiratory
     infection are a heterogeneous family of cells with phagocytic, antigen
     processing and immunomodulatory functions.  Macrophages are important in
     both innate and acquired immunity in the respiratory tract, and have a
     role in lung defense against viruses, bacteria,
     mycobacteria and fungi.  Interactions of pathogens with lung
     macrophages is strongly influenced by sol. immune components including
     complement, collectins and Igs.  Macrophage function can be
     modulated by cytokines, environmental exposures, recent and chronic
     infection including HIV infection, drug therapy and gene
     transfer.
REFERENCE COUNT:         69    THERE ARE 69 CITED REFERENCES AVAILABLE FOR THIS
                               RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT


=> select hit rn l6 1-57;select hit rn l8 1
E1 THROUGH E63 ASSIGNED


NO E#s ASSIGNED

=> fil reg
FILE 'REGISTRY' ENTERED AT 17:06:16 ON 08 JUL 2002
USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.
PLEASE SEE "HELP USAGETERMS" FOR DETAILS.
COPYRIGHT (C) 2002 American Chemical Society (ACS)

STRUCTURE FILE UPDATES:    7 JUL 2002  HIGHEST RN 437604-49-4
DICTIONARY FILE UPDATES:   7 JUL 2002  HIGHEST RN 437604-49-4

TSCA INFORMATION NOW CURRENT THROUGH January 7, 2002

  Please note that search-term pricing does apply when
  conducting SmartSELECT searches.

Crossover limits have been increased.  See HELP CROSSOVER for details.

Calculated physical property data is now available.  See HELP PROPERTIES
for more information.  See STNote 27, Searching Properties in the CAS
Registry File, for complete details:
http://www.cas.org/ONLINE/STN/STNOTES/stnotes27.pdf
```

```
=> d his l14

     (FILE 'HCAPLUS' ENTERED AT 16:58:49 ON 08 JUL 2002)
               SELECT HIT RN L6 1-57

     FILE 'REGISTRY' ENTERED AT 17:06:16 ON 08 JUL 2002
L14           63 S E1-E63

=> d reg l14 1-63
1          RN   372144-62-2  REGISTRY
2          RN   372144-61-1  REGISTRY
3          RN   372144-60-0  REGISTRY
4          RN   372144-59-7  REGISTRY
5          RN   372144-58-6  REGISTRY
6          RN   372144-36-0  REGISTRY
7          RN   372026-68-1  REGISTRY
8          RN   372026-67-0  REGISTRY
9          RN   372026-66-9  REGISTRY
10         RN   372026-65-8  REGISTRY
11         RN   372026-64-7  REGISTRY
12         RN   372025-81-5  REGISTRY
13         RN   372025-80-4  REGISTRY
14         RN   372025-79-1  REGISTRY
15         RN   372025-78-0  REGISTRY
16         RN   372025-77-9  REGISTRY
17         RN   372025-76-8  REGISTRY
18         RN   372025-75-7  REGISTRY
19         RN   372025-74-6  REGISTRY
20         RN   372025-73-5  REGISTRY
21         RN   372025-72-4  REGISTRY
22         RN   372025-71-3  REGISTRY
23         RN   372025-70-2  REGISTRY
24         RN   372025-69-9  REGISTRY
25         RN   372025-68-8  REGISTRY
26         RN   372025-67-7  REGISTRY
27         RN   372025-66-6  REGISTRY
28         RN   372025-65-5  REGISTRY
29         RN   372025-64-4  REGISTRY
30         RN   372025-63-3  REGISTRY
31         RN   372025-61-1  REGISTRY
32         RN   372025-58-6  REGISTRY
33         RN   372025-57-5  REGISTRY
34         RN   372025-56-4  REGISTRY
35         RN   372025-55-3  REGISTRY
36         RN   372025-54-2  REGISTRY
37         RN   372025-50-8  REGISTRY
38         RN   371921-23-2  REGISTRY
39         RN   371921-22-1  REGISTRY
40         RN   371921-21-0  REGISTRY
41         RN   260235-02-7  REGISTRY
42         RN   260235-01-6  REGISTRY
43         RN   260235-00-5  REGISTRY
44         RN   260234-99-9  REGISTRY
45         RN   260234-98-8  REGISTRY
46         RN   260234-97-7  REGISTRY
47         RN   260234-96-6  REGISTRY
48         RN   260234-95-5  REGISTRY
49         RN   260234-94-4  REGISTRY
```

50 RN **260234-89-7** REGISTRY
51 RN **260234-88-6** REGISTRY
52 RN **260234-86-4** REGISTRY
53 RN **252198-24-6** REGISTRY
54 RN **236742-12-4** REGISTRY
55 RN **235094-70-9** REGISTRY
56 RN **226543-52-8** REGISTRY
57 RN **156132-72-8** REGISTRY
58 RN **154837-43-1** REGISTRY
59 RN **154837-42-0** REGISTRY
60 RN **154432-61-8** REGISTRY
61 RN **153058-87-8** REGISTRY
62 RN **142978-86-7** REGISTRY
DR 143107-68-0
63 RN **142978-84-5** REGISTRY

=> d ide can l14 1 5 10 15 20 25 30 35 40 45 50 55 60 63

L14 ANSWER 1 OF 63 REGISTRY COPYRIGHT 2002 ACS
RN **372144-62-2** REGISTRY
CN 1-40-Collectin CL-L2-1 (human) (9CI) (CA INDEX NAME)
OTHER NAMES:
CN 53: PN: WO0181401 SEQID: 53 claimed protein
FS PROTEIN SEQUENCE
MF C183 H302 N48 O52 S2
CI MAN
SR CA
LC STN Files: CA, CAPLUS

*** STRUCTURE DIAGRAM IS NOT AVAILABLE ***
*** USE 'SQD' OR 'SQIDE' FORMATS TO DISPLAY SEQUENCE ***
1 REFERENCES IN FILE CA (1967 TO DATE)
1 REFERENCES IN FILE CAPLUS (1967 TO DATE)

REFERENCE 1: 135:353796

L14 ANSWER 5 OF 63 REGISTRY COPYRIGHT 2002 ACS
RN **372144-58-6** REGISTRY
CN (41-43)-(68-112)-Collectin CL-L2-1 (human) (9CI) (CA INDEX NAME)
OTHER NAMES:
CN 42: PN: WO0181401 SEQID: 42 claimed protein
FS PROTEIN SEQUENCE
MF C194 H319 N61 O66 S
CI MAN
SR CA
LC STN Files: CA, CAPLUS

*** STRUCTURE DIAGRAM IS NOT AVAILABLE ***
*** USE 'SQD' OR 'SQIDE' FORMATS TO DISPLAY SEQUENCE ***
1 REFERENCES IN FILE CA (1967 TO DATE)
1 REFERENCES IN FILE CAPLUS (1967 TO DATE)

REFERENCE 1: 135:353796

L14 ANSWER 10 OF 63 REGISTRY COPYRIGHT 2002 ACS
RN **372026-65-8** REGISTRY
CN DNA (human collectin CL-L2-2 cDNA) (9CI) (CA INDEX NAME)
OTHER NAMES:

CN 46: PN: WO0181401 SEQID: 46 claimed DNA
FS NUCLEIC ACID SEQUENCE
MF Unspecified
CI MAN
SR CA
LC STN Files: CA, CAPLUS

*** STRUCTURE DIAGRAM IS NOT AVAILABLE ***
*** USE 'SQD' OR 'SQIDE' FORMATS TO DISPLAY SEQUENCE ***
1 REFERENCES IN FILE CA (1967 TO DATE)
1 REFERENCES IN FILE CAPLUS (1967 TO DATE)

REFERENCE 1: 135:353796

L14 ANSWER 15 OF 63 REGISTRY COPYRIGHT 2002 ACS
RN **372025-78-0** REGISTRY
CN Collectin CL-L2-1v3 (human) (9CI) (CA INDEX NAME)
OTHER NAMES:
CN 39: PN: WO0181401 SEQID: 39 claimed protein
FS PROTEIN SEQUENCE
MF Unspecified
CI MAN
SR CA
LC STN Files: CA, CAPLUS

*** STRUCTURE DIAGRAM IS NOT AVAILABLE ***
*** USE 'SQD' OR 'SQIDE' FORMATS TO DISPLAY SEQUENCE ***
1 REFERENCES IN FILE CA (1967 TO DATE)
1 REFERENCES IN FILE CAPLUS (1967 TO DATE)

REFERENCE 1: 135:353796

L14 ANSWER 20 OF 63 REGISTRY COPYRIGHT 2002 ACS
RN **372025-73-5** REGISTRY
CN DNA (human collectin CL-L2-1v1 cDNA plus flanks) (9CI) (CA INDEX NAME)
OTHER NAMES:
CN 36: PN: WO0181401 SEQID: 36 claimed DNA
FS NUCLEIC ACID SEQUENCE
MF Unspecified
CI MAN
SR CA
LC STN Files: CA, CAPLUS

*** STRUCTURE DIAGRAM IS NOT AVAILABLE ***
*** USE 'SQD' OR 'SQIDE' FORMATS TO DISPLAY SEQUENCE ***
1 REFERENCES IN FILE CA (1967 TO DATE)
1 REFERENCES IN FILE CAPLUS (1967 TO DATE)

REFERENCE 1: 135:353796

L14 ANSWER 25 OF 63 REGISTRY COPYRIGHT 2002 ACS
RN **372025-68-8** REGISTRY
CN DNA (human collectin CL-L2-2v3 cDNA) (9CI) (CA INDEX NAME)
OTHER NAMES:
CN 57: PN: WO0181401 SEQID: 57 claimed DNA
FS NUCLEIC ACID SEQUENCE
MF Unspecified
CI MAN
SR CA

LC STN Files: CA, CAPLUS

*** STRUCTURE DIAGRAM IS NOT AVAILABLE ***
*** USE 'SQD' OR 'SQIDE' FORMATS TO DISPLAY SEQUENCE ***
1 REFERENCES IN FILE CA (1967 TO DATE)
1 REFERENCES IN FILE CAPLUS (1967 TO DATE)

REFERENCE 1: 135:353796

L14 ANSWER 30 OF 63 REGISTRY COPYRIGHT 2002 ACS
RN **372025-63-3** REGISTRY
CN DNA (human collectin CL-L2-2v1 cDNA) (9CI) (CA INDEX NAME)
OTHER NAMES:
CN 55: PN: WO0181401 SEQID: 55 claimed DNA
FS NUCLEIC ACID SEQUENCE
MF Unspecified
CI MAN
SR CA
LC STN Files: CA, CAPLUS

*** STRUCTURE DIAGRAM IS NOT AVAILABLE ***
*** USE 'SQD' OR 'SQIDE' FORMATS TO DISPLAY SEQUENCE ***
1 REFERENCES IN FILE CA (1967 TO DATE)
1 REFERENCES IN FILE CAPLUS (1967 TO DATE)

REFERENCE 1: 135:353796

L14 ANSWER 35 OF 63 REGISTRY COPYRIGHT 2002 ACS
RN **372025-55-3** REGISTRY
CN DNA (human collectin CL-L2-2 cDNA plus flanks) (9CI) (CA INDEX NAME)
OTHER NAMES:
CN 3: PN: WO0181401 SEQID: 3 claimed DNA
FS NUCLEIC ACID SEQUENCE
MF Unspecified
CI MAN
SR CA
LC STN Files: CA, CAPLUS

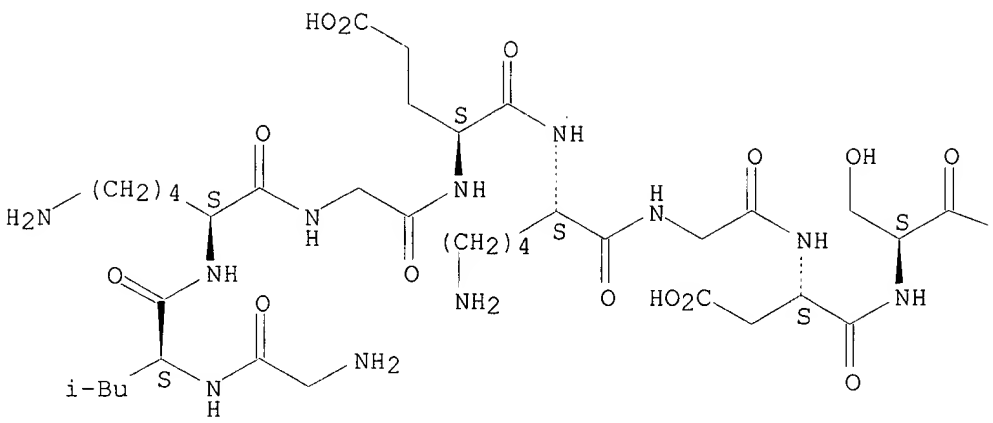
*** STRUCTURE DIAGRAM IS NOT AVAILABLE ***
*** USE 'SQD' OR 'SQIDE' FORMATS TO DISPLAY SEQUENCE ***
1 REFERENCES IN FILE CA (1967 TO DATE)
1 REFERENCES IN FILE CAPLUS (1967 TO DATE)

REFERENCE 1: 135:353796

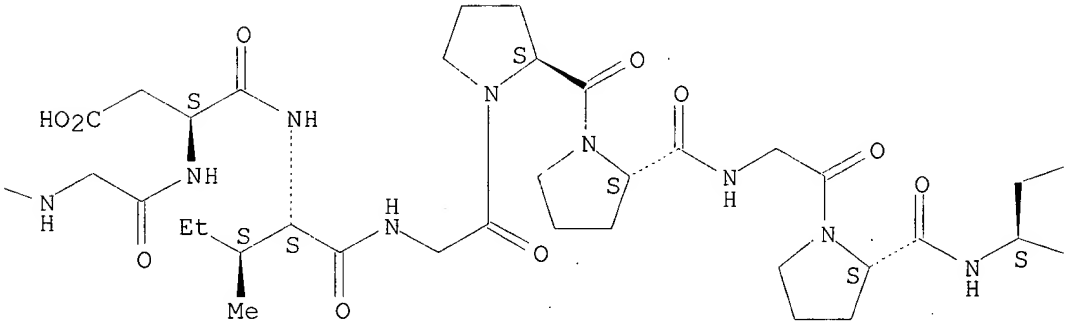
L14 ANSWER 40 OF 63 REGISTRY COPYRIGHT 2002 ACS
RN **371921-21-0** REGISTRY
CN L-Proline, glycyl-L-leucyl-L-lysylglycyl-L-.alpha.-glutamyl-L-lysylglycyl-L-.alpha.-aspartyl-L-serylglycyl-L-.alpha.-aspartyl-L-isoleucylglycyl-L-prolyl-L-prolylglycyl-L-prolyl-L-asparaginylglycyl-L-.alpha.-glutamyl-L-prolylglycyl-L-leucyl- (9CI) (CA INDEX NAME)
OTHER NAMES:
CN (41-43)-(92-112)-Collectin CL-L2-1 (human)
CN 43: PN: WO0181401 SEQID: 43 claimed sequence
FS PROTEIN SEQUENCE; STEREOSEARCH
MF C96 H153 N27 O35
SR CA
LC STN Files: CA, CAPLUS

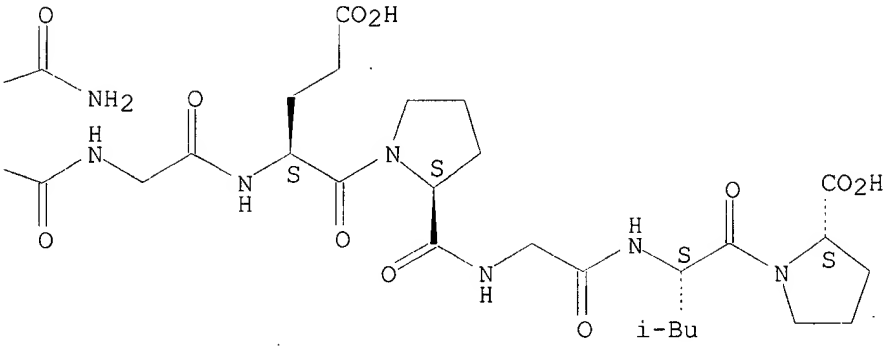
Absolute stereochemistry.

PAGE 1-A



PAGE 1-B





PROPERTY DATA AVAILABLE IN THE 'PROP' FORMAT

1 REFERENCES IN FILE CA (1967 TO DATE)
1 REFERENCES IN FILE CAPLUS (1967 TO DATE)

REFERENCE 1: 135:353796

L14 ANSWER 45 OF 63 REGISTRY COPYRIGHT 2002 ACS
RN **260234-98-8** REGISTRY
CN DNA (human collectin N-terminal fragment-specifying cDNA plus 5'-flank) (9CI) (CA INDEX NAME)
OTHER NAMES:
CN 37: PN: WO0011161 SEQID: 1 claimed DNA
FS NUCLEIC ACID SEQUENCE
MF Unspecified
CI MAN
SR CA
LC STN Files: CA, CAPLUS

*** STRUCTURE DIAGRAM IS NOT AVAILABLE ***
*** USE 'SQD' OR 'SQIDE' FORMATS TO DISPLAY SEQUENCE ***
1 REFERENCES IN FILE CA (1967 TO DATE)
1 REFERENCES IN FILE CAPLUS (1967 TO DATE)

REFERENCE 1: 132:204040

L14 ANSWER 50 OF 63 REGISTRY COPYRIGHT 2002 ACS
RN **260234-89-7** REGISTRY
CN 24-342-Collectin (human) (9CI) (CA INDEX NAME)
OTHER NAMES:
CN 30: PN: WO0011161 SEQID: 2 claimed protein
FS PROTEIN SEQUENCE
MF Unspecified
CI MAN
SR CA
LC STN Files: CA, CAPLUS

*** STRUCTURE DIAGRAM IS NOT AVAILABLE ***
*** USE 'SQD' OR 'SQIDE' FORMATS TO DISPLAY SEQUENCE ***
1 REFERENCES IN FILE CA (1967 TO DATE)
1 REFERENCES IN FILE CAPLUS (1967 TO DATE)

REFERENCE 1: 132:204040

L14 ANSWER 55 OF 63 REGISTRY COPYRIGHT 2002 ACS
RN **235094-70-9** REGISTRY
CN Collectin CL-L1 (collectin liver 1)(human clone HL11-3M) (9CI) (CA INDEX NAME)
OTHER NAMES:
CN Collectin (human liver)
CN Collectin 34 (human clone HL11-3M)
CN GenBank AB002631-derived protein GI 5162875
FS PROTEIN SEQUENCE
MF Unspecified
CI MAN
SR CA
LC STN Files: CA, CAPLUS

*** STRUCTURE DIAGRAM IS NOT AVAILABLE ***
*** USE 'SQD' OR 'SQIDE' FORMATS TO DISPLAY SEQUENCE ***
2 REFERENCES IN FILE CA (1967 TO DATE)
2 REFERENCES IN FILE CAPLUS (1967 TO DATE)

REFERENCE 1: 131:156226

REFERENCE 2: 131:141029

L14 ANSWER 60 OF 63 REGISTRY COPYRIGHT 2002 ACS
RN **154432-61-8** REGISTRY
CN Protein (rat gene WCH-CD water channel-forming reduced) (9CI) (CA INDEX NAME)
OTHER NAMES:
CN Kidney collecting duct water channel(Rat gene WCH-CD)
FS PROTEIN SEQUENCE
MF Unspecified
CI MAN
SR CA
LC STN Files: CA, CAPLUS

*** STRUCTURE DIAGRAM IS NOT AVAILABLE ***
*** USE 'SQD' OR 'SQIDE' FORMATS TO DISPLAY SEQUENCE ***
1 REFERENCES IN FILE CA (1967 TO DATE)
1 REFERENCES IN FILE CAPLUS (1967 TO DATE)

REFERENCE 1: 120:237365

L14 ANSWER 63 OF 63 REGISTRY COPYRIGHT 2002 ACS
RN **142978-84-5** REGISTRY
CN Protein RaRF (mouse clone a10 subunit P28a precursor reduced) (9CI) (CA INDEX NAME)
OTHER NAMES:
CN Collectin (mouse clone .lambda.81 gene Mbl2 precursor reduced)
CN Mannose-binding protein C (mouse clone .lambda.81 gene Mbl2 precursor reduced)
FS PROTEIN SEQUENCE
MF Unspecified

```
CI    MAN
SR    CA
LC    STN Files:    CA, CAPLUS

*** STRUCTURE DIAGRAM IS NOT AVAILABLE ***
*** USE 'SQD' OR 'SQIDE' FORMATS TO DISPLAY SEQUENCE ***
                2 REFERENCES IN FILE CA (1967 TO DATE)
                2 REFERENCES IN FILE CAPLUS (1967 TO DATE)

REFERENCE    1:  123:103916

REFERENCE    2:  117:109915
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 6, 2002, 12:04:10 ; Search time 281.9 Seconds
(without alignments)
9714.362 Million cell updates/sec

Title: US-09-600-932-1
Perfect score: 1595
Sequence: 1 cagcaatgaatggctttgca.....gatttaagaaaaacggagcc 1595

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1595	100.0	1595	20	AAX88323	Human collectin cD
2	980.2	61.5	1016	20	AAZ33973	Human PRO702 nucle
3	980.2	61.5	1016	21	AAC78480	Human PRO702 (UNQ3
4	980.2	61.5	1016	22	AAS45974	Human DNA encoding
5	703.8	44.1	707	22	AAK91268	Human digestive sy
6	246	15.4	1238	21	AAC58385	Human PRO1182 nucl
7	246	15.4	1238	21	AAZ65084	Membrane-bound pro
8	246	15.4	1238	22	AAF44230	Human PRO1182 (UNQ
9	246	15.4	1253	21	AAZ94946	Human carbohydrate
10	246	15.4	1341	24	ABA91171	Human collectin en
11	244.6	15.3	813	24	ABA91201	Human collectin po
12	240.6	15.1	1522	24	ABA91176	Mouse collectin en
13	239	15.0	813	24	ABA91207	Human collectin po
14	235.2	14.7	1139	24	ABA91172	Human collectin en
15	233.8	14.7	735	24	ABA91202	Human collectin po
16	232.2	14.6	990	22	AAH98390	Human EST-derived
17	232.2	14.6	1008	22	AAK51559	Human polynucleoti
18	217.4	13.6	252	21	AAA43156	Human secreted exp
19	211.2	13.2	1269	24	ABA91199	Collectin PCR prim
20	209.8	13.2	741	24	ABA91209	Human collectin po
21	208.2	13.1	1067	24	ABA91174	Human collectin en
22	206.8	13.0	663	24	ABA91205	Human collectin po
23	206.4	12.9	1269	24	ABA91200	Collectin PCR prim
24	205	12.9	741	24	ABA91210	Human collectin po
25	203.4	12.8	1067	24	ABA91175	Human collectin en
26	202	12.7	663	24	ABA91206	Human collectin po
27	192.6	12.1	936	22	AAK51560	Human polynucleoti
28	191.4	12.0	1197	24	ABA91198	Collectin PCR prim
29	190	11.9	669	24	ABA91208	Human collectin po
30	189.4	11.9	995	24	ABA91173	Human collectin en
31	188	11.8	591	24	ABA91204	Human collectin po
32	171.4	10.7	477	24	ABA91203	Human collectin po
33	99	6.2	1373	23	AAS79208	DNA encoding novel
34	94.6	5.9	412	22	AAH33786	Human colon cancer
35	79.2	5.0	318	22	AAK87906	Human digestive sy
36	68.6	4.3	1868	17	AAT39340	DNA sequence for m
37	63.8	4.0	909	22	AAC89872	Human zacrp7 degen
38	63.8	4.0	1954	22	ABA09143	Human HSPP-62 prot
39	63.8	4.0	1954	22	AAK52543	Human polynucleoti
40	63.8	4.0	1954	22	AAK52544	Human polynucleoti
41	63	3.9	6882	20	AAZ10631	Splice variant ZAP
c 42	61	3.8	1510	20	AAX13146	Enterococcus faeca
43	60.4	3.8	10556	22	AAI59459	Human polynucleoti
44	60	3.8	427	22	AAH43028	Nucleotide fragmen
45	60	3.8	2298	22	AAH43022	Nucleotide sequenc

ALIGNMENTS

RESULT 1
AAX88323
ID AAX88323 standard; cDNA; 1595 BP.
XX
AC AAX88323;
XX
DT 30-SEP-1999 (first entry)
XX
DE Human collectin cDNA.
XX
KW Collectin; human; antibacterial; antiviral; treatment; infection; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 6..839
FT /*tag= a
FT /product= "collectin"
XX
PN WO9937767-A1.
XX
PD 29-JUL-1999.
XX
PF 24-JUL-1998; 98WO-JP03328.
XX
PR 23-JAN-1998; 98JP-0011281.
XX
PA (FUSO) FUSO PHARM IND LTD.
XX
PI Wakamiya N;
XX
DR WPI; 1999-458691/38.

```
DR   P-PSDB; AAY25518.
XX
PT   New collectin protein of human origin and DNA encoding it
XX
PS   Claim 2; Page 39-42; 58pp; Japanese.
XX
CC   This invention describes the isolation and characetrisation of a novel
CC   human collectin protein and its encoding polynucleotide. The human
CC   collectin exhibits antibacterial and antiviral activity and can be used
CC   as an agent for the treatment of human bacterial and viral infections.
CC   This sequence encodes the novel human collectin.
XX
SQ   Sequence 1595 BP; 444 A; 322 C; 382 G; 447 T; 0 other;
```

```
  Query Match           100.0%;  Score 1595;  DB 20;  Length 1595;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1595;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 cagcaatgaatggctttgcatccttgcttcgaagaaaccaatttatcctcctggtactat 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 cagcaatgaatggctttgcatccttgcttcgaagaaaccaatttatcctcctggtactat 60

Qy     61 ttcttttgcaaattcagagtctgggtctggatattgatagccgtcctaccgctgaagtct 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ttcttttgcaaattcagagtctgggtctggatattgatagccgtcctaccgctgaagtct 120

Qy    121 gtgccacacacacaatttcaccaggacccaaaggagatgatggtgaaaaaggagatccag 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 gtgccacacacacaatttcaccaggacccaaaggagatgatggtgaaaaaggagatccag 180

Qy    181 gagaagagggaaagcatggcaaagtgggacgcatggggccgaaaggaattaaaggagaac 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 gagaagagggaaagcatggcaaagtgggacgcatggggccgaaaggaattaaaggagaac 240

Qy    241 tgggtgatatgggagatcggggcaatattggcaagactgggcccattgggaagaagggtg 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 tgggtgatatgggagatcggggcaatattggcaagactgggcccattgggaagaagggtg 300

Qy    301 acaaaggggaaaaaggtttgcttggaatacctggagaaaaaggcaaagcaggtactgtct 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 acaaaggggaaaaaggtttgcttggaatacctggagaaaaaggcaaagcaggtactgtct 360

Qy    361 gtgattgtggaagataccggaaatttgttggacaactggatattagtattgcccggctca 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 gtgattgtggaagataccggaaatttgttggacaactggatattagtattgcccggctca 420

Qy    421 agacatctatgaagtttgtcaagaatgtgatagcagggattagggaaactgaagagaaat 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 agacatctatgaagtttgtcaagaatgtgatagcagggattagggaaactgaagagaaat 480

Qy    481 tctactacatcgtgcaggaagagaagaactacagggaatccctaacccactgcaggattc 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 tctactacatcgtgcaggaagagaagaactacagggaatccctaacccactgcaggattc 540

Qy    541 ggggtggaatgctagccatgcccaaggatgaagctgccaacacactcatcgctgactatg 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 ggggtggaatgctagccatgcccaaggatgaagctgccaacacactcatcgctgactatg 600

Qy    601 ttgccaagagtggcttctttcgggtgttcattggcgtgaatgaccttgaaagggagggac 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 ttgccaagagtggcttctttcgggtgttcattggcgtgaatgaccttgaaagggagggac 660

Qy    661 agtacatgttcacagacaacactccactgcagaactatagcaactggaatgaggggggaac 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 agtacatgttcacagacaacactccactgcagaactatagcaactggaatgaggggggaac 720

Qy    721 ccagcgacccctatggtcatgaggactgtgtggagatgctgagctctggcagatggaatg 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 ccagcgacccctatggtcatgaggactgtgtggagatgctgagctctggcagatggaatg 780

Qy    781 acacagagtgccatcttaccatgtactttgtctgtgagttcatcaagaagaaaaagtaac 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 acacagagtgccatcttaccatgtactttgtctgtgagttcatcaagaagaaaaagtaac 840

Qy    841 ttccctcatcctacgtatttgctattttcctgtgaccgtcattacagttattgttatcca 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 ttccctcatcctacgtatttgctattttcctgtgaccgtcattacagttattgttatcca 900

Qy    901 tccttttttttcctgattgtactacatttgatctgagtcaacatagctagaaaatgctaaa 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 tccttttttttcctgattgtactacatttgatctgagtcaacatagctagaaaatgctaaa 960

Qy    961 ctgaggtatggagcctccatcatcatgctcttttgtgatgattttcatattttcacacat 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 ctgaggtatggagcctccatcatcatgctcttttgtgatgattttcatattttcacacat 1020

Qy   1021 ggtatgttattgacccaataactcgccaggttacatgggtcttgagagagaatttttaatt 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 ggtatgttattgacccaataactcgccaggttacatgggtcttgagagagaatttttaatt 1080

Qy   1081 actaattgtgcacgagatagttggttgtctatatgtcaaatgagttgttctcttggtatt 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 actaattgtgcacgagatagttggttgtctatatgtcaaatgagttgttctcttggtatt 1140

Qy   1141 tgctctaccatctctccctagagcactctgtgtctatcccagtggataatttcccagttt 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 tgctctaccatctctccctagagcactctgtgtctatcccagtggataatttcccagttt 1200

Qy   1201 actggtgatgattaggaaggttgttgatggttaggctaacctgccctggcccaaagccag 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 actggtgatgattaggaaggttgttgatggttaggctaacctgccctggcccaaagccag 1260

Qy   1261 acatgtacaagggctttctgtgagcaatgataagatctttgaatccaagatgcccagatg 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 acatgtacaagggctttctgtgagcaatgataagatctttgaatccaagatgcccagatg 1320

Qy   1321 ttttaccagtcacaccctatggccatggctatacttggaagttctccttgttggcacaga 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 ttttaccagtcacaccctatggccatggctatacttggaagttctccttgttggcacaga 1380

Qy   1381 catagaaatgctttaaccccaagcctttatatgggggacttctagctttgtgtcttgttt 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 catagaaatgctttaaccccaagcctttatatgggggacttctagctttgtgtcttgttt 1440

Qy   1441 cagaccatgtggaatgataaatactctttttgtgcttctgatctatcgatttcactaaca 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 cagaccatgtggaatgataaatactctttttgtgcttctgatctatcgatttcactaaca 1500

Qy   1501 tataccaagtaggtgctttgaaccccttttctgtaggctcacaccttaatctcaggcccct 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 tataccaagtaggtgctttgaaccccttttctgtaggctcacaccttaatctcaggcccct 1560

Qy   1561 atatagtcacactttgatttaagaaaaacggagcc 1595
          ||||||||||||||||||||||||||||||||||
Db   1561 atatagtcacactttgatttaagaaaaacggagcc 1595
```

```
RESULT   2
AAZ33973
ID   AAZ33973 standard; cDNA; 1016 BP.
XX
AC   AAZ33973;
XX
DT   07-DEC-1999  (first entry)
XX
DE   Human PRO702 nucleotide sequence.
XX
KW   Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;
KW   probe; blood coagulation disorder; cancer; cellular adhesion disorder;
KW   secreted protein; transmembrane protein; ss.
XX
OS   Homo sapiens.
XX
PN   WO9946281-A2.
XX
```

PD 16-SEP-1999.
XX
PF 08-MAR-1999; 99WO-US05028.
XX
PR 10-MAR-1998; 98US-0077450.
PR 11-MAR-1998; 98US-0077632.
PR 11-MAR-1998; 98US-0077641.
PR 11-MAR-1998; 98US-0077649.
PR 12-MAR-1998; 98US-0077791.
PR 13-MAR-1998; 98US-0078004.
PR 17-MAR-1998; 98US-0040220.
PR 20-MAR-1998; 98US-0078886.
PR 20-MAR-1998; 98US-0078910.
PR 20-MAR-1998; 98US-0078936.
PR 20-MAR-1998; 98US-0078939.
PR 25-MAR-1998; 98US-0079294.
PR 26-MAR-1998; 98US-0079656.
PR 27-MAR-1998; 98US-0079663.
PR 27-MAR-1998; 98US-0079664.
PR 27-MAR-1998; 98US-0079689.
PR 27-MAR-1998; 98US-0079728.
PR 27-MAR-1998; 98US-0079786.
PR 30-MAR-1998; 98US-0079920.
PR 30-MAR-1998; 98US-0079923.
PR 31-MAR-1998; 98US-0080105.
PR 31-MAR-1998; 98US-0080107.
PR 31-MAR-1998; 98US-0080165.
PR 31-MAR-1998; 98US-0080194.
PR 01-APR-1998; 98US-0080327.
PR 01-APR-1998; 98US-0080328.
PR 01-APR-1998; 98US-0080333.
PR 01-APR-1998; 98US-0080334.
PR 08-APR-1998; 98US-0081049.
PR 08-APR-1998; 98US-0081070.
PR 08-APR-1998; 98US-0081071.
PR 09-APR-1998; 98US-0081195.
PR 09-APR-1998; 98US-0081203.
PR 09-APR-1998; 98US-0081229.
PR 15-APR-1998; 98US-0081817.
PR 15-APR-1998; 98US-0081838.
PR 15-APR-1998; 98US-0081952.
PR 15-APR-1998; 98US-0081955.
PR 21-APR-1998; 98US-0082568.
PR 21-APR-1998; 98US-0082569.
PR 22-APR-1998; 98US-0082700.
PR 22-APR-1998; 98US-0082704.
PR 22-APR-1998; 98US-0082804.
PR 23-APR-1998; 98US-0082767.
PR 23-APR-1998; 98US-0082796.
PR 27-APR-1998; 98US-0083336.
PR 28-APR-1998; 98US-0083322.
PR 29-APR-1998; 98US-0083392.
PR 29-APR-1998; 98US-0083495.
PR 29-APR-1998; 98US-0083496.
PR 29-APR-1998; 98US-0083499.
PR 29-APR-1998; 98US-0083500.
PR 29-APR-1998; 98US-0083545.
PR 29-APR-1998; 98US-0083554.
PR 29-APR-1998; 98US-0083558.
PR 29-APR-1998; 98US-0083559.
PR 30-APR-1998; 98US-0083742.
PR 05-MAY-1998; 98US-0084366.
PR 06-MAY-1998; 98US-0084414.
PR 06-MAY-1998; 98US-0084441.
PR 07-MAY-1998; 98US-0084598.
PR 07-MAY-1998; 98US-0084600.
PR 07-MAY-1998; 98US-0084627.
PR 07-MAY-1998; 98US-0084637.
PR 07-MAY-1998; 98US-0084639.
PR 07-MAY-1998; 98US-0084640.
PR 07-MAY-1998; 98US-0084643.
PR 13-MAY-1998; 98US-0085323.
PR 13-MAY-1998; 98US-0085338.
PR 13-MAY-1998; 98US-0085339.
PR 15-MAY-1998; 98US-0085573.
PR 15-MAY-1998; 98US-0085579.
PR 15-MAY-1998; 98US-0085580.
PR 15-MAY-1998; 98US-0085582.
PR 15-MAY-1998; 98US-0085689.
PR 15-MAY-1998; 98US-0085697.
PR 15-MAY-1998; 98US-0085700.
PR 15-MAY-1998; 98US-0085704.
PR 18-MAY-1998; 98US-0086023.
PR 22-MAY-1998; 98US-0086392.
PR 22-MAY-1998; 98US-0086414.
PR 22-MAY-1998; 98US-0086430.
PR 22-MAY-1998; 98US-0086486.
PR 28-MAY-1998; 98US-0087098.
PR 28-MAY-1998; 98US-0087106.
PR 28-MAY-1998; 98US-0087208.
PR 30-JUL-1998; 98US-0094651.
PR 11-SEP-1998; 98US-0100038.
XX
PA (GETH) GENENTECH INC.
XX
PI Wood WI, Goddard A, Gurney A, Yuan J, Baker KP, Chen J;
XX
DR WPI; 1999-551358/46.
DR P-PSDB; AAY41698.
XX
PT New secreted and transmembrane polypeptides and their polynucleotides,
PT useful for treating blood coagulation disorders, cancers and cellular
PT adhesion disorders -
XX
PS Claim 2; Fig 36; 530pp; English.
XX
CC The present invention describes secreted and transmembrane polypeptides
CC and their polynucleotides. The nucleotide sequences are useful as
CC sources of probes, primers, for chromosome mapping, and for generation
CC of antisense sequences. They can also be used to create transgenic
CC animals. The proteins can be used to treat a variety of diseases and
CC disorders, depending on their function. Diseases that may be treated
CC include blood coagulation disorders, cancers and cellular adhesion
CC disorders. They may also be used to raise antibodies. AAZ33891 to
CC AAZ34338, and AAY41685 to AAY41774 represent polynucleotide and
CC polypeptide sequence given in the exemplification of the present
CC invention.
XX
SQ Sequence 1016 BP; 312 A; 197 C; 261 G; 246 T; 0 other;

Query Match 61.5%; Score 980.2; DB 20; Length 1016;
Best Local Similarity 99.7%; Pred. No. 4.9e-278;
Matches 982; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy       1 cagcaatgaatggctttgcatccttgcttcgaagaaaccaatttatcctcctggtactat 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17 cagcaatgaatggctttgcatccttgcttcgaagaaaccaatttatcctcctggtactat 76

Qy      61 ttctttttgcaaattcagagtctgggtctggatattgatagccgtcctaccgctgaagtct 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      77 ttctttttgcaaattcagagtctgggtctggatattgatagccgtcctaccgctgaagtct 136

Qy     121 gtgccacacacacaatttcaccaggacccaaaggagatgatggtgaaaaaggagatccag 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     137 gtgccacacacacaatttcaccaggacccaaaggagatgatggtgaaaaaggagatccag 196

Qy     181 gagaagagggaaagcatggcaaagtgggacgcatggggccgaaaggaattaaaggagaac 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     197 gagaagagggaaagcatggcaaagtgggacgcatggggccgaaaggaattaaaggagaac 256

Qy     241 tgggtgatatgggagatcggggcaatattggcaagactgggcccattgggaagaagggtg 300
           |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db     257 tgggtgatatgggagatcagggcaatattggcaagactgggcccattgggaagaagggtg 316

Qy     301 acaaaggggaaaaaggtttgcttggaatacctggagaaaaaggcaaagcaggtactgtct 360
```

```
Db    317 acaaaggggaaaaaggtttgcttggaatacctggagaaaaaggcaaagcaggtactgtct 376

Qy    361 gtgattgtggaagatacc-ggaaatttgttggacaactggatattagtattgcccggctca 420
          ||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db    377 gtgattgtggaagataccggaaatttgttggacaactggatattagtattgctcggctca 436

Qy    421 agacatctatgaagtttgtcaagaatgtgatagcagggattagggaaactgaagagaaat 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    437 agacatctatgaagtttgtcaagaatgtgatagcagggattagggaaactgaagagaaat 496

Qy    481 tctactacatcgtgcaggaagagaagaagaactacagggaatccctaacccactgcaggattc 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    497 tctactacatcgtgcaggaagagaagaagaactacagggaatccctaacccactgcaggattc 556

Qy    541 ggggtggaatgctagccatgcccaaggatgaagctgccaacacactcatcgctgactatg 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    557 ggggtggaatgctagccatgcccaaggatgaagctgccaacacactcatcgctgactatg 616

Qy    601 ttgccaagagtggcttctttcgggtgttcattggcgtgaatgaccttgaaagggagggac 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    617 ttgccaagagtggcttctttcgggtgttcattggcgtgaatgaccttgaaagggagggac 676

Qy    661 agtacatgttcacagacaacactccactgcagaactatagcaactggaatgaggggaac 720
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    677 agtacatgtccacagacaacactccactgcagaactatagcaactggaatgaggggaac 736

Qy    721 ccagcgacccctatggtcatgaggactgtgtggagatgctgagctctggcagatggaatg 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    737 ccagcgacccctatggtcatgaggactgtgtggagatgctgagctctggcagatggaatg 796

Qy    781 acacagagtgccatcttaccatgtactttgtctgtgagttcatcaagaagaaaaagtaac 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    797 acacagagtgccatcttaccatgtactttgtctgtgagttcatcaagaagaaaaagtaac 856

Qy    841 ttccctcatcctacgtatttgctattttcctgtgaccgtcattacagttattgttatcca 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    857 ttccctcatcctacgtatttgctattttcctgtgaccgtcattacagttattgttatcca 916

Qy    901 tccttttttcctgattgtactacatttgatctgagtcaacatagctagaaaatgctaaa 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    917 tccttttttcctgattgtactacatttgatctgagtcaacatagctagaaaatgctaaa 976

Qy    961 ctgaggtatggagcctccatcatca 985
          |||||||||||||||||||||||||
Db    977 ctgaggtatggagcctccatcatca 1001


RESULT  3
AAC78480
ID   AAC78480 standard; cDNA; 1016 BP.
XX
AC   AAC78480;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Human PRO702 (UNQ366) nucleotide sequence SEQ ID NO:96.
XX
KW   Human; secreted protein; transmembrane protein; PRO; EST; cytostatic;
KW   expressed sequence tag; detection; cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200053756-A2.
XX
PD   14-SEP-2000.
XX
PF   18-FEB-2000; 2000WO-US04341.
XX
PR   08-MAR-1999;   99WO-US05028.
PR   12-MAR-1999;   99US-0123957.
PR   29-MAR-1999;   99US-0126773.
PR   21-APR-1999;   99US-0130232.
PR   28-APR-1999;   99US-0131445.
PR   14-MAY-1999;   99US-0134287.
PR   23-JUN-1999;   99US-0141037.
PR   26-JUL-1999;   99US-0145698.
PR   29-OCT-1999;   99US-0162506.
PR   30-NOV-1999;   99WO-US28313.
PR   02-DEC-1999;   99WO-US28551.
PR   02-DEC-1999;   99WO-US28565.
PR   16-DEC-1999;   99WO-US30095.
PR   30-DEC-1999;   99WO-US31243.
PR   30-DEC-1999;   99WO-US31274.
PR   05-JAN-2000; 2000WO-US00219.
PR   06-JAN-2000; 2000WO-US00277.
PR   06-JAN-2000; 2000WO-US00376.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Ashkenazi AJ,  Baker KP,  Botstein D,  Desnoyers L,  Eaton DL;
PI   Ferrara N,  Filvaroff E,  Fong S,  Gao W,  Gerber H,  Gerritsen ME;
PI   Goddard A,  Godowski PJ,  Grimaldi CJ,  Gurney AL,  Hillan KJ;
PI   Kljavin IJ,  Kuo SS,  Napier MA,  Pan J,  Paoni NF,  Roy MA;
PI   Shelton DL,  Stewart TA,  Tumas D,  Williams PM,  Wood WI;
XX
DR   WPI; 2000-611443/58.
DR   P-PSDB; AAB44254.
XX
PT   Novel PRO polypeptides and polynucleotides used in detection methods,
PT   to target bioactive molecules to specific cells, and to modulate
PT   cellular activities -
XX
PS   Claim 2; Fig 36; 636pp; English.
XX
CC   AAC78458 to AAC78599 represent polynucleotide and EST (expressed
CC   sequence tag) sequences which encode secreted or transmembrane PRO
CC   polypeptides. The PRO polynucleotides and polypeptides have cytostatic
CC   activity. The polynucleotides and polypeptides can be used for detecting
CC   the presence of PRO polypeptides in samples, for linking bioactive
CC   molecules to cells and for modulating biological activities of cells,
CC   using the polypeptides for specific targeting. The polypeptide targeting
CC   can be used to kill the target cells, e.g. for the treatment of cancers.
CC   The polypeptide pairs provide specific targeting of bioactive molecules
CC   to cells. AAC78600 to AAC78987 represent PCR primers and probes used in
CC   the isolation of the PRO polynucleotide sequences.
XX
SQ   Sequence 1016 BP; 312 A; 197 C; 261 G; 246 T; 0 other;


  Query Match           61.5%;  Score 980.2;  DB 21;  Length 1016;
  Best Local Similarity  99.7%;  Pred. No. 4.9e-278;
  Matches  982;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 cagcaatgaatggctttgcatccttgcttcgaagaaaccaatttatcctcctggtactat 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     17 cagcaatgaatggctttgcatccttgcttcgaagaaaccaatttatcctcctggtactat 76

Qy     61 ttcttttgcaaattcagagtctgggtctggatattgatagccgtcctaccgctgaagtct 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     77 ttcttttgcaaattcagagtctgggtctggatattgatagccgtcctaccgctgaagtct 136

Qy    121 gtgccacacacacaatttcaccaggacccaaaggagatgatggtgaaaaaggagatccag 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    137 gtgccacacacacaatttcaccaggacccaaaggagatgatggtgaaaaaggagatccag 196

Qy    181 gagaagagggaaagcatggcaaagtgggacgcatggggccgaaaggaattaaaggagaac 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    197 gagaagagggaaagcatggcaaagtgggacgcatggggccgaaaggaattaaaggagaac 256

Qy    241 tgggtgatatgggagatcggggcaatattggcaagactgggcccattgggaagaagggtg 300
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    257 tgggtgatatgggagatcagggcaatattggcaagactgggcccattgggaagaagggtg 316

Qy    301 acaaaggggaaaaaggtttgcttggaatacctggagaaaaaggcaaagcaggtactgtct 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    317 acaaaggggaaaaaggtttgcttggaatacctggagaaaaaggcaaagcaggtactgtct 376

Qy    361 gtgattgtggaagataccggaaatttgttggacaactggatattagtattgcccggctca 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db    377 gtgattgtggaagataccggaaatttgttggacaactggatattagtattgctcggctca 436

Qy    421 agacatctatgaagtttgtcaagaatgtgatagcagggattagggaaactgaagagaaat 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    437 agacatctatgaagtttgtcaagaatgtgatagcagggattagggaaactgaagagaaat 496

Qy    481 tctactacatcgtgcaggaagagaagaactacagggaatccctaacccactgcaggattc 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||L
Db    497 tctactacatcgtgcaggaagagaagaactacagggaatccctaacccactgcaggattc 556

Qy    541 ggggtggaatgctagccatgcccaaggatgaagctgccaacacactcatcgctgactatg 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    557 ggggtggaatgctagccatgcccaaggatgaagctgccaacacactcatcgctgactatg 616

Qy    601 ttgccaagagtggcttctttcgggtgttcattggcgtgaatgaccttgaaagggagggac 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    617 ttgccaagagtggcttctttcgggtgttcattggcgtgaatgaccttgaaagggagggac 676

Qy    661 agtacatgttcacagacaacactccactgcagaactatagcaactggaatgaggggaac 720
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    677 agtacatgtccacagacaacactccactgcagaactatagcaactggaatgaggggaac 736

Qy    721 ccagcgacccctatggtcatgaggactgtgtggagatgctgagctctggcagatggaatg 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    737 ccagcgacccctatggtcatgaggactgtgtggagatgctgagctctggcagatggaatg 796

Qy    781 acacagagtgccatcttaccatgtactttgtctgtgagttcatcaagaagaaaaagtaac 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    797 acacagagtgccatcttaccatgtactttgtctgtgagttcatcaagaagaaaaagtaac 856

Qy    841 ttccctcatcctacgtatttgctattttcctgtgaccgtcattacagttattgttatcca 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    857 ttccctcatcctacgtatttgctattttcctgtgaccgtcattacagttattgttatcca 916

Qy    901 tccttttttttcctgattgtactacatttgatctgagtcaacatagctagaaaatgctaaa 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    917 tccttttttttcctgattgtactacatttgatctgagtcaacatagctagaaaatgctaaa 976

Qy    961 ctgaggtatggagcctccatcatca 985
          |||||||||||||||||||||||||
Db    977 ctgaggtatggagcctccatcatca 1001
```

```
RESULT   4
AAS45974
ID   AAS45974 standard; cDNA; 1016 BP.
XX
AC   AAS45974;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Human DNA encoding PRO polypeptide sequence #50.
XX
KW   PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep; ss;
KW   dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;
KW   blood; chondrocyte cell; cell proliferation; cell differentiation; colon;
KW   adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder;
KW   PCR primer.
XX
OS   Homo sapiens.
XX
PN   WO200168848-A2.
XX
PD   20-SEP-2001.
XX
PF   28-FEB-2001; 2001WO-US06520.
XX
PR   01-MAR-2000; 2000WO-US05601.
PR   02-MAR-2000; 2000WO-US05841.
PR   03-MAR-2000; 2000US-187202P.
PR   06-MAR-2000; 2000US-186968P.
PR   14-MAR-2000; 2000US-189320P.
PR   14-MAR-2000; 2000US-189328P.
PR   15-MAR-2000; 2000WO-US06884.
PR   21-MAR-2000; 2000US-190828P.
PR   21-MAR-2000; 2000US-191007P.
PR   21-MAR-2000; 2000US-191048P.
PR   21-MAR-2000; 2000US-191314P.
PR   28-MAR-2000; 2000US-192655P.
PR   29-MAR-2000; 2000US-193032P.
PR   29-MAR-2000; 2000US-193053P.
PR   30-MAR-2000; 2000WO-US08439.
PR   04-APR-2000; 2000US-194449P.
PR   04-APR-2000; 2000US-194647P.
PR   11-APR-2000; 2000US-195975P.
PR   11-APR-2000; 2000US-196000P.
PR   11-APR-2000; 2000US-196187P.
PR   11-APR-2000; 2000US-196690P.
PR   11-APR-2000; 2000US-196820P.
PR   18-APR-2000; 2000US-198121P.
PR   18-APR-2000; 2000US-198585P.
PR   25-APR-2000; 2000US-199397P.
PR   25-APR-2000; 2000US-199550P.
PR   25-APR-2000; 2000US-199654P.
PR   03-MAY-2000; 2000US-201516P.
PR   17-MAY-2000; 2000WO-US13705.
PR   22-MAY-2000; 2000WO-US14042.
PR   30-MAY-2000; 2000WO-US14941.
PR   02-JUN-2000; 2000WO-US15264.
PR   05-JUN-2000; 2000US-209832P.
PR   28-JUL-2000; 2000WO-US20710.
PR   22-AUG-2000; 2000US-0644848.
PR   24-AUG-2000; 2000WO-US23328.
PR   08-NOV-2000; 2000WO-US30952.
PR   01-DEC-2000; 2000WO-US32678.
PR   20-DEC-2000; 2000WO-US34956.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Baker KP,  Chen J,  Desnoyers L,  Goddard A,  Godowski PJ,  Gurney AL;
PI   Pan J,  Smith V,  Watanabe CK,  Wood WI,  Zhang Z;
XX
DR   WPI; 2001-602746/68.
DR   P-PSDB; AAU29073.
XX
PT   Novel nucleic acids encoding PRO polypeptides, used to diagnose the
PT   presence of tumours, such as prostate and breast tumours, in mammals and
PT   to screen for modulators of the compounds -
XX
PS   Claim 2; Fig 99; 774pp; English.
XX
CC   Sequences AAS45925-AAS46231 represent DNA molecules encoding and PCR
CC   primers for PRO polypeptides of the invention. The sequences of the
CC   invention can be used to detect the presence of a tumour in a mammal by
CC   comparing the level of expression of a PRO polypeptide in a test sample
CC   of cells from the animal and a control sample of normal cells, whereby a
CC   higher level of expression in the test sample indicates the presence of a
CC   tumour in the mammal. Mammals include dogs, cats, cattle, horses, sheep,
CC   pigs, goats and rabbits but are preferably human. The polypeptides can be
CC   used to stimulate tumour necrosis factor (TNF) alpha release from human
CC   blood, when contacted with it. A specific polypeptide can be used to
CC   stimulate the proliferation or differentiation of chondrocyte cells. The
CC   PRO proteins can be used to determine the presence of tumours and also
CC   susceptibility to tumour development, particularly adrenal, lung, colon,
CC   breast, prostate, rectal, cervical, or liver tumours, in mammalian
CC   subjects. The oligonucleotide probes specific for the PRO nucleic acids
CC   can be used for genetic analysis of individuals with genetic disorders.
XX
SQ   Sequence 1016 BP; 312 A; 197 C; 261 G; 246 T; 0 other;

  Query Match             61.5%;  Score 980.2;  DB 22;  Length 1016;
```

```
  Best Local Similarity   99.7%;  Pred. No. 4.9e-278;
  Matches  982;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      1 cagcaatgaatggctttgcatccttgcttcgaagaaaccaatttatcctcctggtactat 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     17 cagcaatgaatggctttgcatccttgcttcgaagaaaccaatttatcctcctggtactat 76

Qy     61 ttcttttgcaaattcagagtctgggtctggatattgatagccgtcctaccgctgaagtct 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     77 ttcttttgcaaattcagagtctgggtctggatattgatagccgtcctaccgctgaagtct 136

Qy    121 gtgccacacacacaatttcaccaggacccaaaggagatgatggtgaaaaaggagatccag 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    137 gtgccacacacacaatttcaccaggacccaaaggagatgatggtgaaaaaggagatccag 196

Qy    181 gagaagagggaaagcatggcaaagtgggacgcatggggccgaaaggaattaaaggagaac 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    197 gagaagagggaaagcatggcaaagtgggacgcatggggccgaaaggaattaaaggagaac 256

Qy    241 tgggtgatatgggagatcggggcaatattggcaagactgggcccattgggaagaagggtg 300
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    257 tgggtgatatgggagatcagggcaatattggcaagactgggcccattgggaagaagggtg 316

Qy    301 acaaaggggaaaaaggtttgcttggaatacctggagaaaaaggcaaagcaggtactgtct 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    317 acaaaggggaaaaaggtttgcttggaatacctggagaaaaaggcaaagcaggtactgtct 376

Qy    361 gtgattgtggaagataccggaaatttgttggacaactggatattagtattgcccggctca 420
          |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db    377 gtgattgtggaagataccggaaatttgttggacaactggatattagtattgctcggctca 436

Qy    421 agacatctatgaagtttgtcaagaatgtgatagcagggattagggaaactgaagagaaat 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    437 agacatctatgaagtttgtcaagaatgtgatagcagggattagggaaactgaagagaaat 496

Qy    481 tctactacatcgtgcaggaagagaagaactacagggaatccctaacccactgcaggattc 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    497 tctactacatcgtgcaggaagagaagaactacagggaatccctaacccactgcaggattc 556

Qy    541 ggggtggaatgctagccatgcccaaggatgaagctgccaacacactcatcgctgactatg 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    557 ggggtggaatgctagccatgcccaaggatgaagctgccaacacactcatcgctgactatg 616

Qy    601 ttgccaagagtggcttctttcgggtgttcattggcgtgaatgaccttgaaagggagggac 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    617 ttgccaagagtggcttctttcgggtgttcattggcgtgaatgaccttgaaagggagggac 676

Qy    661 agtacatgttcacagacaacactccactgcagaactatagcaactggaatgaggggaac 720
          ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    677 agtacatgtccacagacaacactccactgcagaactatagcaactggaatgaggggaac 736

Qy    721 ccagcgacccctatggtcatgaggactgtgtggagatgctgagctctggcagatggaatg 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    737 ccagcgacccctatggtcatgaggactgtgtggagatgctgagctctggcagatggaatg 796

Qy    781 acacagagtgccatcttaccatgtactttgtctgtgagttcatcaagaagaaaaagtaac 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    797 acacagagtgccatcttaccatgtactttgtctgtgagttcatcaagaagaaaaagtaac 856

Qy    841 ttccctcatcctacgtatttgctattttcctgtgaccgtcattacagttattgttatcca 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    857 ttccctcatcctacgtatttgctattttcctgtgaccgtcattacagttattgttatcca 916

Qy    901 tccttttttttcctgattgtactacatttgatctgagtcaacatagctagaaaatgctaaa 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    917 tccttttttttcctgattgtactacatttgatctgagtcaacatagctagaaaatgctaaa 976

Qy    961 ctgaggtatggagcctccatcatca 985
          |||||||||||||||||||||||||
Db    977 ctgaggtatggagcctccatcatca 1001


RESULT   5
```

```
AAK91268
ID   AAK91268 standard; DNA; 707 BP.
XX
AC   AAK91268;
XX
DT   05-NOV-2001  (first entry)
XX
DE   Human digestive system antigen genomic sequence SEQ ID NO: 4844.
XX
KW   Human; digestive system antigen; gene therapy; cancer; appendicitis;
KW   ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW   digestive system disorder; Meckel's diverticulum; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155314-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01324.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
```

PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-502630/55.
XX
PT Polynucleotides encoding digestive system antigens, useful for
PT diagnosing, treating, preventing and/or prognozing disorders of the
PT digestive system, particularly cancer and cancer metastases -
XX
PS Disclosure; SEQ ID NO 4844; 986pp; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human digestive system antigens. These can be used in the
CC diagnosis, treatment and prevention of digestive system disorders,
CC including cancer, Meckel's diverticulum, bacterial or parasitic
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC ulcerative colitis. The present sequence is a genomic DNA fragment
CC encoding a digestive system antigen of the invention.
XX
SQ Sequence 707 BP; 185 A; 145 C; 154 G; 223 T; 0 other;

Query Match 44.1%; Score 703.8; DB 22; Length 707;
Best Local Similarity 99.7%; Pred. No. 9.8e-197;
Matches 705; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy   624 gtgttcattggcgtgaatgaccttgaaagggagggacagtacatgttcacagacaacact 683
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 gtgttcattggcgtgaatgaccttgaaagggagggacagtacatgttcacagacaacact 60

Qy   684 ccactgcagaactatagcaactggaatgaggggggaacccagcgacccctatggtcatgag 743
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 ccactgcagaactatagcaactggaatgaggggggaacccagcgacccctatggtcatgag 120

Qy   744 gactgtgtggagatgctgagctctggcagatggaatgacacagagtgccatcttaccatg 803
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 gactgtgtggagatgctgagctctggcagatggaatgacacagagtgccatcttaccatg 180

Qy   804 tactttgtctgtgagttcatcaagaagaaaaagtaacttccctcatcctacgtatttgct 863
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 tactttgtctgtgagttcatcaagaagaaaaagtaacttccctcatcctacgtatttgct 240

Qy   864 attttcctgtgaccgtcattacagttattgttatccatccttttttttcctgattgtacta 923
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 attttcctgtgaccgtcattacagttattgttatccatccttttttttcctgattgtacta 300

Qy   924 catttgatctgagtcaacatagctagaaaatgctaaactgaggtatggagcctccatcat 983
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 catttgatctgagtcaacatagctagaaaatgctaaactgaggtatggagcctccatcat 360

Qy   984 catgctcttttgtgatgattttcatattttcacacatggtatgttattgacccaataact 1043
         |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db   361 catgctcttttgtgatgattttcatattttcacacatggtatattattgacccaataact 420

Qy  1044 cgccaggttacatgggtcttgagagagaattttaattactaattgtgcacgagatagttg 1103
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 cgccaggttacatgggtcttgagagagaattttaattactaattgtgcacgagatagttg 480

Qy  1104 gttgtctatatgtcaaatgagttgttctcttggtatttgctctaccatctctccctagag 1163
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    481 gttgtctatatgtcaaatgagttgttctcttggtatttgctctaccatctctccctagag 540

Qy   1164 cactctgtgtctatcccagtggataatttcccagtttactggtgatgattaggaaggttg 1223
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 cactctgtgtctatcccagtggataatttcccagtttactggtgatgattaggaaggttg 600

Qy   1224 ttgatggttaggctaacctgccctggcccaaagccagacatgtacaagggctttctgtga 1283
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 ttgatggttaggctaacctgccctggcccaaagccagacatgtacaagggctttctgtga 660

Qy   1284 gcaatgataagatctttgaatccaagatgcccagatgttttaccagt 1330
          |||||||||||||||||||||||||||||||| ||||||||||
Db    661 gcaatgataagatctttgaatccaagatgcccagatcttttaccagt 707
```

RESULT 6
AAC58385
ID AAC58385 standard; cDNA; 1238 BP.
XX
AC AAC58385;
XX
DT 29-JAN-2001 (first entry)
XX
DE Human PRO1182 nucleotide sequence SEQ ID NO:50.
XX
KW Human; tumour; diagnosis; neoplastic disease; neoplastic cell growth;
KW proliferation; tumourigenesis; identification; cancer; cytostatic;
KW nootropic; neuroprotective; antiinflammatory; immunosuppressive;
KW immunostimulant; antiangiogenic; leukaemia; lymphoid malignancy;
KW neuronal disorder; glial disorder; astrocytal disorder; angiogenic;
KW hypothalamic disorder; glandular disorder; macrophagal disorder;
KW epithelial disorder; stromal disorder; blastocoelic disorder;
KW inflammatory disorder; immunologic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200053755-A2.
XX
PD 14-SEP-2000.
XX
PF 06-JAN-2000; 2000WO-US00376.
XX
PR 08-MAR-1999; 99WO-US05028.
PR 02-JUN-1999; 99WO-US12252.
PR 23-JUN-1999; 99US-0141037.
PR 07-JUL-1999; 99US-0143048.
PR 26-JUL-1999; 99US-0145698.
PR 30-NOV-1999; 99WO-US28313.
PR 20-DEC-1999; 99WO-US30911.
PR 05-JAN-2000; 2000WO-US00219.
XX
PA (GETH) GENENTECH INC.
XX
PI Ashkenazi AJ, Baker KP, Goddard A, Gurney AL, Hillan KJ, Roy MA;
PI Watanabe CK, Wood WI;
XX
DR WPI; 2000-572270/53.
DR P-PSDB; AAB24075.
XX
PT Thirty PRO polynucleotides encoding PRO polypeptides, useful in the
PT treatment, diagnosis and prevention of cancer -
XX
PS Claim 50; Fig 37; 286pp; English.
XX
CC The present invention describes an isolated antibody that binds to
CC one of the human PRO proteins designated PRO212, PRO290, PRO341, PRO535,
CC PRO619, PRO717, PRO809, PRO830, PRO848, PRO943, PRO1005, PRO1009,
CC PRO1025, PRO1030, PRO1097, PRO1107, PRO1111, PRO1153, PRO1182, PRO1184,
CC PRO1187, PRO1281, PRO23, PRO39, PRO834, PRO1317, PRO1710, PRO2094,
CC PRO2145 OR PRO2198. PRO antagonists can be used to inhibit tumour cell
CC growth. The PRO polypeptides and nucleotides are useful in the
CC treatment, diagnosis and prevention of cancer. The antibodies and other
CC anti-tumour compounds maybe used to treat various conditions, including
CC those characterised by overexpression and/or activation of the amplified
CC PRO genes. Exemplary conditions or disorders to be treated with such
CC antibodies and other compounds include benign or malignant tumours
CC (e.g., renal, liver, kidney, bladder, breast, gastric, ovarian,
CC colorectal, prostate, pancreatic, lung, vulva, thyroid, hepatic
CC carcinomas, sarcomas, glioblastomas, and various head and neck tumours),
CC leukaemias and lymphoid malignancies, other disorders such as neuronal,
CC glial, astrocytal, hypothalamic and other glandular, macrophagal,
CC epithelial, stromal and blastocoelic disorders, and inflammatory,
CC angiogenic and immunologic disorders. AAC58242 to AAC58366 represent PCR
CC primers and hybridisation probes used in the isolation of the human PRO
CC sequences. AAC58367 to AAC58396 and AAB24057 to AAB24089 represent human
CC PRO polynucleotide and protein sequences given in the exemplification of
CC the present invention.
XX
SQ Sequence 1238 BP; 293 A; 321 C; 389 G; 235 T; 0 other;

Query Match 15.4%; Score 246; DB 21; Length 1238;
Best Local Similarity 58.4%; Pred. No. 8.4e-62;
Matches 429; Conservative 0; Mismatches 305; Indels 0; Gaps 0;

```
Qy    106 ctaccgctgaagtctgtgccacacacacaatttcaccaggacccaaaggagatgatggtg 165
          ||  || ||| | |||  |    || |   |    || || | |||||| |||| ||| |
Db    149 ctggcgatgacgcctgctctgtgcagatcctcgtccctggcctcaaaggggatgcgggag 208

Qy    166 aaaaaggagatccaggagaagagggaaagcatggcaaagtgggacgcatggggccgaaag 225
          | || |||||   ||| |     ||| || ||| | ||| || | || |||    ||||
Db    209 agaagggagacaaaggcgcccccggacggcctggaagagtcggccccacgggagaaaaag 268

Qy    226 gaattaaaggagaactgggtgatatgggagatcggggcaatattggcaagactgggccca 285
          ||   |  || ||    ||  | |  ||   |  |||   |  ||| || | ||| ||||
Db    269 gagacatgggggacaaaggacagaaaggcagtgtgggtcgtcatggaaaaattggtccca 328

Qy    286 ttgggaagaagggtgacaaaggggaaaaaggtttgcttggaatacctggagaaaaaggca 345
          ||||    || ||||| ||||| ||  ' |||    | |||   |||||    || ||
Db    329 ttggctctaaaggtgagaaaggagattccggtgacataggacccccctggtcctaatggag 388

Qy    346 aagcaggtactgtctgtgattgtggaagataccggaaatttgttggacaactggatatta 405
          || ||||       ||||| ||  |       || ||     | ||  |  |||| |
Db    389 aaccaggcctcccatgtgagtgcagccagctgcgcaaggccatcggggagatggacaacc 448

Qy    406 gtattgcccggctcaagacatctatgaagtttgtcaagaatgtgatagcagggattaggg 465
          |  | | ||| |  |       | |||||  |||||||||   | || ||  |  | |
Db    449 aggtctctcagctgaccagcgagctcaagttcatcaagaatgctgtcgccggtgtgcgcg 508

Qy    466 aaactgaagagaaattctactacatcgtgcaggaagagaagaactacagggaatccctaa 525
          | || ||    ||  |||||    | ||| |||| ||||||  ||||  |||  |||
Db    509 agacggagagcaagatctacctgctggtgaaggaggagaagcgctacgcggacgcccagc 568

Qy    526 cccactgcaggattcggggtggaatgctagccatgcccaaggatgaagctgccaacacac 585
          |||| |   || || || | |||   ||||||||||| || ||||||||    |
Db    569 tgtcctgccagggccgcgggggcacgctgagcatgcccaaggacgaggctgccaatggcc 628

Qy    586 tcatcgctgactatgttgccaagagtggcttctttcgggtgttcattggcgtgaatgacc 645
          | || || |  ||  | ||  |    ||| |    || || ||||| ||| | || ||||
Db    629 tgatggccgcatacctggcgcaagccggcctggcccgtgtcttcatcggcatcaacgacc 688

Qy    646 ttgaaagggagggacagtacatgttcacagacaacactccactgcagaactatagcaact 705
          | || | ||||||    | | ||| | | ||| || | ||  ||| || ||  | ||| |
Db    689 tggagaaggagggcgccttcgtgtactctgaccactcccccatgcggaccttcaacaagt 748

Qy    706 ggaatgaggggggaacccagcgacccctatggtcatgaggactgtgtggagatgctgagct 765
          ||      || || |||| | |  |||| |   | |||||||| |||||||||  ||  ||
Db    749 ggcgcagcggtgagcccaacaatgcctacgacgaggaggactgcgtggagatggtggcct 808

Qy    766 ctggcagatggaatgacacagagtgccatcttaccatgtactttgtctgtgagttcatca 825
          | ||| | ||||| |||   |  |||||    |||||||||||  | ||||||||   ||
Db    809 cgggcggctggaacgacgtggcctgccacaccaccatgtacttcatgtgtgagtttgaca 868

Qy    826 agaagaaaaagtaa 839
          || |||| | || |
Db    869 aggagaacatgtga 882
```

RESULT 7
AAZ65084
ID AAZ65084 standard; cDNA; 1238 BP.
XX
AC AAZ65084;
XX
DT 05-APR-2000 (first entry)
XX
DE Membrane-bound protein PRO1182 encoding cDNA.
XX
KW Membrane-bound polypeptide; PRO polypeptide; LDL receptor; TIE ligand;
KW pharmaceutical; receptor immunoadhesin; gene mapping; ss.
XX
OS Homo sapiens.
XX
PN WO9963088-A2.
XX
PD 09-DEC-1999.
XX
PF 02-JUN-1999; 99WO-US12252.
XX
PR 02-JUN-1998; 98US-0087607.
PR 02-JUN-1998; 98US-0087609.
PR 02-JUN-1998; 98US-0087759.
PR 03-JUN-1998; 98US-0087827.
PR 04-JUN-1998; 98US-0088021.
PR 04-JUN-1998; 98US-0088025.
PR 04-JUN-1998; 98US-0088028.
PR 04-JUN-1998; 98US-0088029.
PR 04-JUN-1998; 98US-0088030.
PR 04-JUN-1998; 98US-0088033.
PR 04-JUN-1998; 98US-0088326.
PR 05-JUN-1998; 98US-0088167.
PR 05-JUN-1998; 98US-0088202.
PR 05-JUN-1998; 98US-0088212.
PR 05-JUN-1998; 98US-0088217.
PR 09-JUN-1998; 98US-0088655.
PR 10-JUN-1998; 98US-0088722.
PR 10-JUN-1998; 98US-0088730.
PR 10-JUN-1998; 98US-0088734.
PR 10-JUN-1998; 98US-0088738.
PR 10-JUN-1998; 98US-0088740.
PR 10-JUN-1998; 98US-0088741.
PR 10-JUN-1998; 98US-0088742.
PR 10-JUN-1998; 98US-0088810.
PR 10-JUN-1998; 98US-0088811.
PR 10-JUN-1998; 98US-0088824.
PR 10-JUN-1998; 98US-0088825.
PR 10-JUN-1998; 98US-0088826.
PR 11-JUN-1998; 98US-0088858.
PR 11-JUN-1998; 98US-0088861.
PR 11-JUN-1998; 98US-0088863.
PR 11-JUN-1998; 98US-0088876.
PR 12-JUN-1998; 98US-0089090.
PR 12-JUN-1998; 98US-0089105.
PR 16-JUN-1998; 98US-0089440.
PR 16-JUN-1998; 98US-0089512.
PR 16-JUN-1998; 98US-0089514.
PR 17-JUN-1998; 98US-0089532.
PR 17-JUN-1998; 98US-0089538.
PR 17-JUN-1998; 98US-0089598.
PR 17-JUN-1998; 98US-0089599.
PR 17-JUN-1998; 98US-0089600.
PR 17-JUN-1998; 98US-0089653.
PR 18-JUN-1998; 98US-0089801.
PR 18-JUN-1998; 98US-0089907.
PR 18-JUN-1998; 98US-0089908.
PR 19-JUN-1998; 98US-0089947.
PR 19-JUN-1998; 98US-0089948.
PR 19-JUN-1998; 98US-0089952.
PR 22-JUN-1998; 98US-0090246.
PR 22-JUN-1998; 98US-0090252.
PR 22-JUN-1998; 98US-0090254.
PR 23-JUN-1998; 98US-0090349.
PR 23-JUN-1998; 98US-0090355.
PR 24-JUN-1998; 98US-0090429.
PR 24-JUN-1998; 98US-0090431.
PR 24-JUN-1998; 98US-0090435.
PR 24-JUN-1998; 98US-0090444.
PR 24-JUN-1998; 98US-0090445.
PR 24-JUN-1998; 98US-0090461.
PR 24-JUN-1998; 98US-0090472.
PR 24-JUN-1998; 98US-0090535.
PR 24-JUN-1998; 98US-0090538.
PR 24-JUN-1998; 98US-0090540.
PR 24-JUN-1998; 98US-0090557.
PR 25-JUN-1998; 98US-0090676.
PR 25-JUN-1998; 98US-0090678.
PR 25-JUN-1998; 98US-0090688.
PR 25-JUN-1998; 98US-0090690.
PR 25-JUN-1998; 98US-0090691.
PR 25-JUN-1998; 98US-0090694.
PR 25-JUN-1998; 98US-0090695.
PR 25-JUN-1998; 98US-0090696.
PR 26-JUN-1998; 98US-0090862.
PR 26-JUN-1998; 98US-0090863.
PR 01-JUL-1998; 98US-0091358.
PR 01-JUL-1998; 98US-0091360.
PR 01-JUL-1998; 98US-0091544.
PR 02-JUL-1998; 98US-0091478.
PR 02-JUL-1998; 98US-0091486.
PR 02-JUL-1998; 98US-0091519.
PR 02-JUL-1998; 98US-0091626.
PR 02-JUL-1998; 98US-0091628.
PR 02-JUL-1998; 98US-0091633.
PR 02-JUL-1998; 98US-0091646.
PR 02-JUL-1998; 98US-0091673.
PR 07-JUL-1998; 98US-0091978.
PR 07-JUL-1998; 98US-0091982.
PR 09-JUL-1998; 98US-0092182.
PR 10-JUL-1998; 98US-0092472.
PR 20-JUL-1998; 98US-0093339.
PR 30-JUL-1998; 98US-0094651.
PR 04-AUG-1998; 98US-0095282.
PR 04-AUG-1998; 98US-0095285.
PR 04-AUG-1998; 98US-0095301.
PR 04-AUG-1998; 98US-0095302.
PR 04-AUG-1998; 98US-0095318.
PR 04-AUG-1998; 98US-0095321.
PR 04-AUG-1998; 98US-0095325.
PR 10-AUG-1998; 98US-0095916.
PR 10-AUG-1998; 98US-0095929.
PR 10-AUG-1998; 98US-0096012.
PR 11-AUG-1998; 98US-0096143.
PR 11-AUG-1998; 98US-0096146.
PR 12-AUG-1998; 98US-0096329.
PR 17-AUG-1998; 98US-0096757.
PR 17-AUG-1998; 98US-0096766.
PR 17-AUG-1998; 98US-0096768.
PR 17-AUG-1998; 98US-0096773.
PR 17-AUG-1998; 98US-0096791.
PR 17-AUG-1998; 98US-0096867.
PR 17-AUG-1998; 98US-0096891.
PR 17-AUG-1998; 98US-0096894.
PR 17-AUG-1998; 98US-0096895.
PR 17-AUG-1998; 98US-0096897.
PR 18-AUG-1998; 98US-0096949.
PR 18-AUG-1998; 98US-0096950.
PR 18-AUG-1998; 98US-0096959.
PR 18-AUG-1998; 98US-0096960.
PR 18-AUG-1998; 98US-0097022.
PR 19-AUG-1998; 98US-0097141.
PR 20-AUG-1998; 98US-0097218.
PR 24-AUG-1998; 98US-0097661.

```
PR   26-AUG-1998;   98US-0097951.
PR   26-AUG-1998;   98US-0097952.
PR   26-AUG-1998;   98US-0097954.
PR   26-AUG-1998;   98US-0097955.
PR   26-AUG-1998;   98US-0097971.
PR   26-AUG-1998;   98US-0097974.
PR   26-AUG-1998;   98US-0097978.
PR   26-AUG-1998;   98US-0097979.
PR   26-AUG-1998;   98US-0097986.
PR   26-AUG-1998;   98US-0098014.
PR   31-AUG-1998;   98US-0098525.
PR   16-SEP-1998;   98US-0100634.
PR   12-JAN-1999;   99US-0115565.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Baker K,  Chen J,  Goddard A,  Gurney AL,  Smith V,  Watanabe CK;
PI   Wood WI,  Yuan J;
XX
DR   WPI; 2000-072883/06.
DR   P-PSDB; AAY66738.
XX
PT   Membrane-bound proteins and related nucleotide sequences -
XX
PS   Claim 2; Fig 251; 822pp; English.
XX
CC   The invention provides membrane-bound PRO polypeptides and
CC   polynucleotides encoding them. The PRO sequences of the invention were
CC   identified based on extracellular domain homology screening. The PRO
CC   sequences have homology with proteins including LDL receptors, TIE
CC   ligands and various enzymes. The membrane-bound proteins and receptor
CC   molecules are useful as pharmaceutical and diagnostic agents. Receptor
CC   immunoadhesins, for instance, can be used as therapeutic agents to block
CC   receptor-ligand interactions. The membrane-bound proteins can also be
CC   employed for screening of potential peptide or small molecule inhibitors
CC   of the relevant receptor/ligand interaction. The PRO encoding sequences
CC   are useful as hybridization probes, in chromosome and gene mapping and in
CC   the generation of antisense RNA and DNA. PRO nucleic acid sequences
CC   will also be useful for the preparation of PRO polypeptides, especially
CC   by recombinant techniques.
XX
SQ   Sequence 1238 BP; 293 A; 321 C; 389 G; 235 T; 0 other;

  Query Match           15.4%;  Score 246;  DB 21;  Length 1238;
  Best Local Similarity  58.4%;  Pred. No. 8.4e-62;
  Matches  429;  Conservative    0;  Mismatches  305;  Indels    0;  Gaps    0;

Qy    106 ctaccgctgaagtctgtgccacacacacaatttcaccaggacccaaaggagatgatggtg 165
          ||  || ||| | |||  |    || |   |    || || | |||||| ||||  || |
Db    149 ctggcgatgacgcctgctctgtgcagatcctcgtccctggcctcaaaggggatgcgggag 208

Qy    166 aaaaaggagatccaggagaagagggaaagcatggcaaagtgggacgcatggggccgaaag 225
          | || |||||   ||| |     |||  || ||| | ||| || | || |||    ||||
Db    209 agaagggagacaaaggcgcccccggacggcctggaagagtcggccccacgggagaaaaag 268

Qy    226 gaattaaaggagaactgggtgatatgggagatcggggcaatattggcaagactgggccca 285
          ||   |  || ||    ||  | |  ||   |  |||   |  ||| || | ||| ||||
Db    269 gagacatgggggacaaaggacagaaaggcagtgtgggtcgtcatggaaaaattggtccca 328

Qy    286 ttgggaagaagggtgacaaaggggaaaaaggtttgcttggaatacctggagaaaaaggca 345
          ||||    || ||||| ||||| ||    |||    | |||   |||||    || ||
Db    329 ttggctctaaaggtgagaaaggagattccggtgacataggaccccctggtcctaatggag 388

Qy    346 aagcaggtactgtctgtgattgtggaagataccggaaatttgttggacaactggatatta 405
          || ||||       ||||| ||  |        || ||     | ||  |  |||| |
Db    389 aaccaggcctcccatgtgagtgcagccagctgcgcaaggccatcggggagatggacaacc 448

Qy    406 gtattgcccggctcaagacatctatgaagtttgtcaagaatgtgatagcagggattaggg 465
             | | | ||| |  |      | |||||  |||||||||   | || ||  |  | |
Db    449 aggtctctcagctgaccagcgagctcaagttcatcaagaatgctgtcgccggtgtgcgcg 508

Qy    466 aaactgaagagaaattctactacatcgtgcaggaagagaagaactacagggaatccctaa 525
```

```
          | || ||    ||  |||||    | ||| |||| ||||||  ||||  |||  |||
Db    509 agacggagagcaagatctacctgctggtgaaggaggagaagcgctacgcggacgcccagc 568

Qy    526 cccactgcaggattcggggtggaatgctagccatgcccaaggatgaagctgccaacacac 585
              ||||  |   || || || | |||   ||||||||||| || ||||||||    |
Db    569 tgtcctgccagggccgcgggggcacgctgagcatgcccaaggacgaggctgccaatggcc 628

Qy    586 tcatcgctgactatgttgccaagagtggcttctttcgggtgttcattggcgtgaatgacc 645
          | || || |  ||  | ||  |    ||| |    || || ||||| ||| | || ||||
Db    629 tgatggccgcatacctggcgcaagccggcctggcccgtgtcttcatcggcatcaacgacc 688

Qy    646 ttgaaagggagggacagtacatgttcacagacaacactccactgcagaactatagcaact 705
          | || | ||||||    | | ||| | | ||| || | ||  ||| || ||  | ||| |
Db    689 tggagaaggagggcgccttcgtgtactctgaccactcccccatgcggaccttcaacaagt 748

Qy    706 ggaatgaggggggaacccagcgacccctatggtcatgaggactgtgtggagatgctgagct 765
          ||      || || |||| | |  |||| |   | |||||||| |||||||| ||  ||
Db    749 ggcgcagcggtgagcccaacaatgcctacgacgaggaggactgcgtggagatggtggcct 808

Qy    766 ctggcagatggaatgacacagagtgccatcttaccatgtactttgtctgtgagttcatca 825
          | ||| | ||||| |||   |  |||||    |||||||||||  | ||||||||   ||
Db    809 cgggcggctggaacgacgtggcctgccacaccaccatgtacttcatgtgtgagtttgaca 868

Qy    826 agaagaaaaagtaa 839
          || |||| | || |
Db    869 aggagaacatgtga 882
```

```
RESULT   8
AAF44230
ID   AAF44230 standard; cDNA; 1238 BP.
XX
AC   AAF44230;
XX
DT   02-APR-2001  (first entry)
XX
DE   Human PRO1182 (UNQ596) nucleotide sequence SEQ ID NO:356.
XX
KW   Human; secreted and transmembrane protein; PRO; cytostatic;
KW   cell death; cancer; chromosomal mapping; gene mapping; tissue typing;
KW   diagnostic assay; ss.
XX
OS   Homo sapiens.
XX
PN   WO200073454-A1.
XX
PD   07-DEC-2000.
XX
PF   30-MAR-2000; 2000WO-US08439.
XX
PR   02-JUN-1999;   99WO-US12252.
PR   23-JUN-1999;   99US-0141037.
PR   07-JUL-1999;   99US-0143048.
PR   20-JUL-1999;   99US-0144758.
PR   26-JUL-1999;   99US-0145698.
PR   28-JUL-1999;   99US-0146222.
PR   17-AUG-1999;   99US-0149396.
PR   15-SEP-1999;   99WO-US21090.
PR   15-SEP-1999;   99WO-US21547.
PR   08-OCT-1999;   99US-0158663.
PR   30-NOV-1999;   99WO-US28313.
PR   01-DEC-1999;   99WO-US28301.
PR   16-DEC-1999;   99WO-US30095.
PR   20-DEC-1999;   99WO-US30911.
PR   05-JAN-2000;  2000WO-US00219.
PR   06-JAN-2000;  2000WO-US00376.
PR   11-FEB-2000;  2000WO-US03565.
PR   18-FEB-2000;  2000WO-US04341.
PR   22-FEB-2000;  2000WO-US04414.
PR   24-FEB-2000;  2000WO-US04914.
PR   24-FEB-2000;  2000WO-US05004.
PR   02-MAR-2000;  2000WO-US05841.
PR   15-MAR-2000;  2000WO-US06884.
```

```
PR   20-MAR-2000; 2000WO-US07377.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Ashkenazi AJ,  Baker KP,  Botstein D,  Desnoyers L,  Eaton DL;
PI   Ferrara N,  Fong S,  Gerber H,  Gerritsen ME,  Goddard A,  Godowski PJ;
PI   Grimaldi CJ,  Gurney AL,  Kljavin IJ,  Napier MA,  Pan J,  Paoni NF;
PI   Roy MA,  Stewart TA,  Tumas D,  Watanabe CK,  Williams PM,  Wood WI;
PI   Zhang Z;
XX
DR   WPI; 2001-032160/04.
DR   P-PSDB; AAB65261.
XX
PT   PRO polynucleotides used to produce polypeptides used to target
PT   bioactive molecules such as toxins, radiolabels or antibodies, to
PT   specific cells, to cause targeted cell death -
XX
PS   Claim 2; Fig 251; 935pp; English.
XX
CC   The present invention describes human secreted and transmembrane PRO
CC   proteins. The PRO proteins have cytostatic activity. The PRO proteins
CC   can be used for targeted delivery of bioactive molecules, such as
CC   toxins, radiolabels or antibodies, that cause cell death. PRO nucleotide
CC   sequences, and their fragments, can be used as hybridisation probes, in
CC   chromosomal and gene mapping, and in the generation of anti-sense RNA
CC   and DNA. They may also be used to produce transgenic animals which are
CC   used to develop and screen therapeutically useful reagents. The PRO
CC   nucleotide and protein sequence can be used for tissue typing and in
CC   treating cancer. Anti-PRO antibodies can be used in diagnostic assays.
CC   AAF44270 to AAF44470 represent PCR primers and hybridisation probes used
CC   in the isolation of human PRO sequences. AAF44087 to AAF44269 and
CC   AAB65154 to AAB65300 represent human PRO polynucleotide and protein
CC   sequences given in the exemplification of the present invention.
XX
SQ   Sequence 1238 BP; 293 A; 321 C; 389 G; 235 T; 0 other;


  Query Match             15.4%;  Score 246;  DB 22;  Length 1238;
  Best Local Similarity   58.4%;  Pred. No. 8.4e-62;
  Matches  429;  Conservative    0;  Mismatches  305;  Indels    0;  Gaps    0;

Qy    106 ctaccgctgaagtctgtgccacacacacaatttcaccaggacccaaaggagatgatggtg 165
          ||  || ||| | |||  |    || |   |    || || | |||||| ||||  || |
Db    149 ctggcgatgacgcctgctctgtgcagatcctcgtccctggcctcaaaggggatgcgggag 208

Qy    166 aaaaaggagatccaggagaagagggaaagcatggcaaagtgggacgcatggggccgaaag 225
          | || |||||   ||| |     |||  || ||| | ||| || | || |||    ||||
Db    209 agaagggagacaaaggcgcccccggacggcctggaagagtcggccccacgggagaaaaag 268

Qy    226 gaattaaaggagaactgggtgatatgggagatcggggcaatattggcaagactgggccca 285
          ||   |  || ||    ||  | |  ||   |  |||   |  ||| || | ||| ||||
Db    269 gagacatgggggacaaaggacagaaaggcagtgtgggtcgtcatggaaaaattggtccca 328

Qy    286 ttgggaagaagggtgacaaaggggaaaaaggtttgcttggaatacctggagaaaaaggca 345
          ||||    || ||||| ||||| ||    |||    | |||   |||||    || ||
Db    329 ttggctctaaaggtgagaaaggagattccggtgacataggacccccctggtcctaatggag 388

Qy    346 aagcaggtactgtctgtgattgtggaagataccggaaatttgttggacaactggatatta 405
          || ||||       ||||| ||  |       || ||     | ||  |  |||| |
Db    389 aaccaggcctcccatgtgagtgcagccagctgcgcaaggccatcggggagatggacaacc 448

Qy    406 gtattgcccggctcaagacatctatgaagtttgtcaagaatgtgatagcagggattaggg 465
            |  | | ||| |  |       | |||||  |||||||||   | || ||  |  | |
Db    449 aggtctctcagctgaccagcgagctcaagttcatcaagaatgctgtcgccggtgtgcgcg 508

Qy    466 aaactgaagagaaattctactacatcgtgcaggaagagaagaactacagggaatccctaa 525
          | || ||    ||  |||||    | ||| |||| ||||||  ||||  |||  |||
Db    509 agacggagagcaagatctacctgctggtgaaggaggagaagcgctacgcggacgcccagc 568

Qy    526 cccactgcaggattcggggtggaatgctagccatgcccaaggatgaagctgccaacacac 585
              ||||  |   || || || | |||   ||||||||||||| || |||||||||    |
Db    569 tgtcctgccagggccgcgggggcacgctgagcatgcccaaggacgaggctgccaatggcc 628

Qy    586 tcatcgctgactatgttgccaagagtggcttctttcgggtgttcattggcgtgaatgacc 645
          | || || |  ||  | ||  |    ||| |    || || ||||| ||| | || ||||
Db    629 tgatggccgcatacctggcgcaagccggcctggcccgtgtcttcatcggcatcaacgacc 688

Qy    646 ttgaaagggagggacagtacatgttcacagacaacactccactgcagaactatagcaact 705
          | || | ||||||    | | ||| | | ||| || | ||  ||| || ||  | ||| |
Db    689 tggagaaggagggcgccttcgtgtactctgaccactccccatgcggaccttcaacaagt 748

Qy    706 ggaatgaggggggaacccagcgacccctatggtcatgaggactgtgtggagatgctgagct 765
          ||       || || |||| | |  |||| |   | |||||||| |||||||||  ||  ||
Db    749 ggcgcagcggtgagcccaacaatgcctacgacgaggaggactgcgtggagatggtggcct 808

Qy    766 ctggcagatggaatgacacagagtgccatcttaccatgtactttgtctgtgagttcatca 825
          | ||| | ||||| |||   |  |||||    ||||||||||  | |||||||   ||
Db    809 cgggcggctggaacgacgtggcctgccacaccaccatgtacttcatgtgtgagtttgaca 868

Qy    826 agaagaaaaagtaa 839
          || |||| | || |
Db    869 aggagaacatgtga 882


RESULT   9
AAZ94946
ID   AAZ94946 standard; cDNA; 1253 BP.
XX
AC   AAZ94946;
XX
DT   01-AUG-2000  (first entry)
XX
DE   Human carbohydrate-associated protein CRBAP-6 cDNA.
XX
KW   CRBAP-6; carbohydrate-associated protein 6; human;
KW   autoimmune disorder; inflammation; gastrointestinal disorder;
KW   infection; reproductive disorder; neurological disorder;
KW   eye disorder; cell proliferation; cancer; diagnosis; gene therapy;
KW   ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             90..905
FT                   /*tag=  a
FT   sig_peptide     90..149
FT                   /*tag=  b
FT   mat_peptide     150..902
FT                   /*tag=  c
XX
PN   WO200018922-A2.
XX
PD   06-APR-2000.
XX
PF   29-SEP-1999;   99WO-US22685.
XX
PR   01-OCT-1998;   98US-0164785.
PR   06-OCT-1998;   98US-0167179.
PR   13-NOV-1998;   98US-0191838.
PR   03-DEC-1998;   98US-0205656.
PR   03-DEC-1998;   98US-2223333.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Au-Young J,  Lal P,  Bandman O,  Reddy R,  Baughn MR,  Yue H;
PI   Hillman JL;
XX
DR   WPI; 2000-317516/27.
DR   P-PSDB; AAY79510.
XX
PT   Novel carbohydrate-associated proteins used for the prevention and
PT   treatment of autoimmune/inflammatory disorders of e.g. the
PT   gastrointestinal and reproductive systems -
XX
PS   Claim 9; Page 94; 104pp; English.
XX
```

CC The present sequence is that of cDNA coding for a novel human
CC carbohydrate-associated protein, termed CRBAP-6 (see AAY79510).
CC The cDNA (Incyte clone 2821011) was initially identified in
CC adrenal tumour cDNA library ADRETUT06. CRBAP-6 has chemical and
CC structural similarity with bovine lung surfactant protein D (32%
CC identity). CRBAP-6 is expressed in the liver, kidney, ovary, gut,
CC adrenal gland and secretory epithelium. The invention provides
CC CRBAP-1 to -7 polynucleotides (see AAZ94941-48) and polypeptides (see
CC AAY79505-11), as well as expression vectors, host cells, antibodies,
CC agonists and antagonists. These are used in the diagnosis,
CC treatment or prevention of disorders associated with CRBAP
CC expression, especially autoimmune or inflammatory disorders,
CC gastrointestinal disorders, infectious disorders, reproductive
CC disorders, neurological disorders, eye disorders and cell
CC proliferative disorders, including cancer. CRBAP polynucleotides
CC are useful sources of probes and primers which can be used to
CC detect CRBAP in a sample from a patient. They may also be
CC administered as part of a gene therapy regime.
XX
SQ Sequence 1253 BP; 287 A; 328 C; 400 G; 238 T; 0 other;

Query Match 15.4%; Score 246; DB 21; Length 1253;
Best Local Similarity 58.4%; Pred. No. 8.4e-62;
Matches 429; Conservative 0; Mismatches 305; Indels 0; Gaps 0;

```
Qy    106 ctaccgctgaagtctgtgccacacacaatttcaccaggacccaaaggagatgatggtg 165
          ||  || ||| | |||  |    ||  |   |    || || | ||||||  ||||  || |
Db    172 ctggcgatgacgcctgctctgtgcagatcctcgtccctggcctcaaaggggatgcgggag 231

Qy    166 aaaaaggagatccaggagaagagggaaagcatggcaaagtgggacgcatggggccgaaag 225
          | || |||||   ||| |     |||  || ||| | ||| || | || |||    ||||
Db    232 agaagggagacaaaggcgcccccggacggcctggaagagtcggccccacgggagaaaaag 291

Qy    226 gaattaaaggagaactgggtgatatgggagatcggggcaatattggcaagactgggccca 285
          ||   |  || ||    ||  | |  ||   |  |||   |  ||| || | ||| ||||
Db    292 gagacatgggggacaaaggacagaaaggcagtgtgggtcgtcatggaaaaattggtccca 351

Qy    286 ttgggaagaagggtgacaaaggggaaaaaggtttgcttggaatacctggagaaaaaggca 345
          ||||    || ||||| ||||| ||    |||    | |||   |||||    || ||
Db    352 ttggctctaaaggtgagaaaggagattccggtgacataggacccccctggtcctaatggag 411

Qy    346 aagcaggtactgtctgtgattgtggaagataccggaaatttgttggacaactggatatta 405
          || ||||       ||||| ||  |        || ||     | ||  |  |||| |
Db    412 aaccaggcctcccatgtgagtgcagccagctgcgcaaggccatcggggagatggacaacc 471

Qy    406 gtattgcccggctcaagacatctatgaagtttgtcaagaatgtgatagcagggattaggg 465
          |  | | ||| |  |      | |||||  |||||||||   | || ||  |  | |
Db    472 aggtctctcagctgaccagcgagctcaagttcatcaagaatgctgtcgccggtgtgcgcg 531

Qy    466 aaactgaagagaaattctactacatcgtgcaggaagagaagaactacagggaatccctaa 525
          | || ||    ||  |||||    | ||| |||| ||||||  ||||  |||  |||
Db    532 agacggagagcaagatctacctgctggtgaaggaggagaagcgctacgcggacgcccagc 591

Qy    526 cccactgcaggattcggggtggaatgctagccatgcccaaggatgaagctgccaacacac 585
            ||||  |   || || || | |||   ||||||||||| || ||||||||    |
Db    592 tgtcctgccaggccgcgggggcacgctgagcatgcccaaggacgaggctgccaatggcc 651

Qy    586 tcatcgctgactatgttgccaagagtggcttctttcgggtgttcattggcgtgaatgacc 645
          | || || |  ||  | ||  |    ||| |    || || ||||| ||| | || ||||
Db    652 tgatggccgcatacctggcgcaagccggcctggcccgtgtcttcatcggcatcaacgacc 711

Qy    646 ttgaaagggagggacagtacatgttcacagacaacactccactgcagaactatagcaact 705
          | || | ||||||    | | ||| | | ||| || | ||  ||| || ||  | ||| |
Db    712 tggagaaggagggcgccttcgtgtactctgaccactccccatgcggaccttcaacaagt 771

Qy    706 ggaatgaggggggaacccagcgacccctatggtcatgaggactgtgtggagatgctgagct 765
          ||      || || |||| | |  |||| |   | |||||||| |||||||||| ||  ||
Db    772 ggcgcagcggtgagcccaacaatgcctacgacgaggaggactgcgtggagatggtggcct 831

Qy    766 ctggcagatggaatgacacagagtgccatcttaccatgtactttgtctgtgagttcatca 825
          | ||| | ||||| |||   |  |||||    ||||||||||  | ||||||||   ||
Dh    832 cgggcggctggaacgacgtggcctgccacaccaccatgtacttcatgtgtgagtttgaca 891
```

Qy 826 agaagaaaaagtaa 839
|| |||| | || |
Db 892 aggagaacatgtga 905

RESULT 10
ABA91171
ID ABA91171 standard; cDNA; 1341 BP.
XX
AC ABA91171;
XX
DT 19-FEB-2002 (first entry)
XX
DE Human collectin encoding polynucleotide SEQ ID NO 1.
XX
KW Human; collectin; CL-L2-1; CL-L2-2; mouse; antibacterial; virucide;
KW protein therapy; infection; ss.
XX
OS Homo sapiens.
XX
PN WO200181401-A1.
XX
PD 01-NOV-2001.
XX
PF 23-APR-2001; 2001WO-JP03468.
XX
PR 21-APR-2000; 2000JP-0120358.
XX
PA (FUSO) FUSO PHARM IND LTD.
XX
PI Wakamiya N, Keshi H, Ohtani K, Sakamoto T, Kishi Y;
XX
DR WPI; 2002-055345/07.
DR P-PSDB; ABB56407.
XX
PT New collectin family proteins, designated CL-L2-1 and CL-L2-2,
PT expressed in kidney and for treatment and prevention of bacterial and
PT viral infections
XX
PS Claim 2; Page 88-90; 134pp; Japanese.
XX
CC The invention relates to human collectin family proteins (CL-L2-1 and
CC CL-L2-2, sequences given in the specification, ABB56407-ABB56411 and
CC ABB56414-ABB56416), their derivatives and fragments and a related
CC collectin (CL-L2) of mouse origin (ABB56412) and polynucleotides encoding
CC all or part of the proteins. The proteins have antibacterial and virucide
CC activity and are used for protein therapy and treatment, prevention and
CC diagnosis of bacterial and viral infections. The present sequence is that
CC of a collectin polynucleotide of the invention.
XX
SQ Sequence 1341 BP; 277 A; 375 C; 449 G; 240 T; 0 other;

Query Match 15.4%; Score 246; DB 24; Length 1341;
Best Local Similarity 58.4%; Pred. No. 8.8e-62;
Matches 429; Conservative 0; Mismatches 305; Indels 0; Gaps 0;

```
Qy    106 ctaccgctgaagtctgtgccacacacaatttcaccaggacccaaaggagatgatggtg 165
          ||  || ||| | |||  |    ||  |   |    || || | ||||||  ||||  || |
Db    347 ctggcgatgacgcctgctctgtgcagatcctcgtccctggcctcaaaggggatgcgggag 406

Qy    166 aaaaaggagatccaggagaagagggaaagcatggcaaagtgggacgcatggggccgaaag 225
          | || |||||   ||| |     |||  || ||| | ||| || | || |||    ||||
Db    407 agaagggagacaaaggcgcccccggacggcctggaagagtcggccccacgggagaaaaag 466

Qy    226 gaattaaaggagaactgggtgatatgggagatcggggcaatattggcaagactgggccca 285
          ||   |  || ||    ||  | |  ||   |  |||   |  ||| || | ||| ||||
Db    467 gagacatgggggacaaaggacagaaaggcagtgtgggtcgtcatggaaaaattggtccca 526

Qy    286 ttgggaagaagggtgacaaaggggaaaaaggtttgcttggaatacctggagaaaaaggca 345
          ||||    || ||||| ||||| ||    |||    | |||   |||||    || ||
Db    527 ttggctctaaaggtgagaaaggagattccggtgacataggacccccctggtcctaatggag 586
```

```
Qy    346 aagcaggtactgtctgtgattgtggaagataccggaaatttgttggacaactggatatta 405
          || ||||       ||||| ||  |       || ||     | ||  |  |||| |
Db    587 aaccaggcctcccatgtgagtgcagccagctgcgcaaggccatcggggagatggacaacc 646

Qy    406 gtattgcccggctcaagacatctatgaagtttgtcaagaatgtgatagcagggattaggg 465
          |  | | ||| |  |       | |||||  |||||||||   | || ||  |  | |
Db    647 aggtctctcagctgaccagcgagctcaagttcatcaagaatgctgtcgccggtgtgcgcg 706

Qy    466 aaactgaagagaaattctactacatcgtgcaggaagagaagaactacagggaatccctaa 525
          | || ||    ||  |||||    | ||| |||| ||||||  ||||  |||  |||
Db    707 agacggagagcaagatctacctgctggtgaaggaggagaagcgctacgcggacgcccagc 766

Qy    526 cccactgcaggattcggggtggaatgctagccatgcccaaggatgaagctgccaacacac 585
              ||||  |   || || || | |||   ||||||||||| || ||||||||    |
Db    767 tgtcctgccagggccgcgggggcacgctgagcatgcccaaggacgaggctgccaatggcc 826

Qy    586 tcatcgctgactatgttgccaagagtggcttctttcgggtgttcattggcgtgaatgacc 645
          | || || |  ||  | ||  |    ||| |    || || |||| ||| | || ||||
Db    827 tgatggccgcatacctggcgcaagccggcctggcccgtgtcttcatcggcatcaacgacc 886

Qy    646 ttgaaagggagggacagtacatgttcacagacaacactccactgcagaactatagcaact 705
          | || | ||||||    | | ||| | | ||| || | ||  ||| || ||  | ||| |
Db    887 tggagaaggagggcgccttcgtgtactctgaccactcccccatgcggaccttcaacaagt 946

Qy    706 ggaatgaggggggaacccagcgacccctatggtcatgaggactgtgtggagatgctgagct 765
          ||      || || |||| | |  |||| |   | ||||||| |||||||| ||  ||
Db    947 ggcgcagcggtgagcccaacaatgcctacgacgaggaggactgcgtggagatggtggcct 1006

Qy    766 ctggcagatggaatgacacagagtgccatcttaccatgtactttgtctgtgagttcatca 825
          | ||| | ||||| |||   |  |||||    ||||||||||  | ||||||||   ||
Db   1007 cgggcggctggaacgacgtggcctgccacaccaccatgtacttcatgtgtgagtttgaca 1066

Qy    826 agaagaaaaagtaa 839
          || |||| | || |
Db   1067 aggagaacatgtga 1080
```

RESULT 11
ABA91201
ID ABA91201 standard; DNA; 813 BP.
XX
AC ABA91201;
XX
DT 19-FEB-2002 (first entry)
XX
DE Human collectin polynucleotide SEQ ID NO 45.
XX
KW Human; collectin; CL-L2-1; CL-L2-2; mouse; antibacterial; virucide;
KW protein therapy; infection; ds.
XX
OS Homo sapiens.
XX
PN WO200181401-A1.
XX
PD 01-NOV-2001.
XX
PF 23-APR-2001; 2001WO-JP03468.
XX
PR 21-APR-2000; 2000JP-0120358.
XX
PA (FUSO) FUSO PHARM IND LTD.
XX
PI Wakamiya N, Keshi H, Ohtani K, Sakamoto T, Kishi Y;
XX
DR WPI; 2002-055345/07.
XX
PT New collectin family proteins, designated CL-L2-1 and CL-L2-2,
PT expressed in kidney and for treatment and prevention of bacterial and
PT viral infections -
XX
PS Claim 2; Page 121; 134pp; Japanese.
XX
CC The invention relates to human collectin family proteins (CL-L2-1 and
CC CL-L2-2, sequences given in the specification, ABB56407-ABB56411 and
CC ABB56414-ABB56416), their derivatives and fragments and a related
CC collectin (CL-L2) of mouse origin (ABB56412) and polynucleotides encoding
CC all or part of the proteins. The proteins have antibacterial and virucide
CC activity and are used for protein therapy and treatment, prevention and
CC diagnosis of bacterial and viral infections. The present sequence is that
CC of a collectin polynucleotide of the invention.
XX
SQ Sequence 813 BP; 183 A; 225 C; 271 G; 134 T; 0 other;

Query Match 15.3%; Score 244.6; DB 24; Length 813;
Best Local Similarity 58.4%; Pred. No. 1.7e-61;
Matches 427; Conservative 0; Mismatches 304; Indels 0; Gaps 0;

```
Qy    106 ctaccgctgaagtctgtgccacacacacaatttcaccaggacccaaaggagatgatggtg 165
          ||  || ||| | |||  |    || |   |    || || | |||||| ||||  || |
Db     83 ctggcgatgacgcctgctctgtgcagatcctcgtccctggcctcaaaggggatgcgggag 142

Qy    166 aaaaaggagatccaggagaagagggaaagcatggcaaagtgggacgcatggggccgaaag 225
          | || |||||   ||| |     |||  || ||| | ||| || | || |||    ||||
Db    143 agaagggagacaaaggcgcccccggacggcctggaagagtcggccccacgggagaaaaag 202

Qy    226 gaattaaaggagaactgggtgatatgggagatcggggcaatattggcaagactgggccca 285
          ||   | || ||    ||  | |  ||   |  |||   |  ||| || | ||| ||||
Db    203 gagacatgggggacaaaggacagaaaggcagtgtgggtcgtcatggaaaaattggtccca 262

Qy    286 ttgggaagaagggtgacaaaggggaaaaaggtttgcttggaatacctggagaaaaaggca 345
          ||||   || ||||| ||||| ||    |||    | |||   |||||    || ||
Db    263 ttggctctaaaggtgagaaaggagattccggtgacataggacccccctggtcctaatggag 322

Qy    346 aagcaggtactgtctgtgattgtggaagataccggaaatttgttggacaactggatatta 405
          || ||||       ||||| ||  |       || ||     | ||  |  |||| |
Db    323 aaccaggcctcccatgtgagtgcagccagctgcgcaaggccatcggggagatggacaacc 382

Qy    406 gtattgcccggctcaagacatctatgaagtttgtcaagaatgtgatagcagggattaggg 465
          |  | | ||| |  |       | |||||  |||||||||   | || ||  |  | |
Db    383 aggtctctcagctgaccagcgagctcaagttcatcaagaatgctgtcgccggtgtgcgcg 442

Qy    466 aaactgaagagaaattctactacatcgtgcaggaagagaagaactacagggaatccctaa 525
          | || ||    ||  |||||    | ||| |||| ||||||  ||||  |||  |||
Db    443 agacggagagcaagatctacctgctggtgaaggaggagaagcgctacgcggacgcccagc 502

Qy    526 cccactgcaggattcggggtggaatgctagccatgcccaaggatgaagctgccaacacac 585
              ||||  |   || || || | |||   ||||||||||| || ||||||||    |
Db    503 tgtcctgccagggccgcgggggcacgctgagcatgcccaaggacgaggctgccaatggcc 562

Qy    586 tcatcgctgactatgttgccaagagtggcttctttcgggtgttcattggcgtgaatgacc 645
          | || || |  ||  | ||  |    ||| |    || || |||| ||| | || ||||
Db    563 tgatggccgcatacctggcgcaagccggcctggcccgtgtcttcatcggcatcaacgacc 622

Qy    646 ttgaaagggagggacagtacatgttcacagacaacactccactgcagaactatagcaact 705
          | || | ||||||    | | ||| | | ||| || | ||  ||| || ||  | ||| |
Db    623 tggagaaggagggcgccttcgtgtactctgaccactcccccatgcggaccttcaacaagt 682

Qy    706 ggaatgaggggggaacccagcgacccctatggtcatgaggactgtgtggagatgctgagct 765
          ||      || || |||| | |  |||| |   | ||||||| |||||||| ||  ||
Db    683 ggcgcagcggtgagcccaacaatgcctacgacgaggaggactgcgtggagatggtggcct 742

Qy    766 ctggcagatggaatgacacagagtgccatcttaccatgtactttgtctgtgagttcatca 825
          | ||| | ||||| |||   |  |||||    ||||||||||  | ||||||||   ||
Db    743 cgggcggctggaacgacgtggcctgccacaccaccatgtacttcatgtgtgagtttgaca 802

Qy    826 agaagaaaaag 836
          || |||| | |
Db    803 aggagaacatg 813
```

RESULT 12
ABA91176
ID ABA91176 standard; cDNA; 1522 BP.
XX

```
AC   ABA91176;
XX
DT   19-FEB-2002  (first entry)
XX
DE   Mouse collectin encoding polynucleotide SEQ ID NO 12.
XX
KW   Human; collectin; CL-L2-1; CL-L2-2; mouse; antibacterial; virucide;
KW   protein therapy; infection; ss.
XX
OS   Mus musculus.
XX
PN   WO200181401-A1.
XX
PD   01-NOV-2001.
XX
PF   23-APR-2001; 2001WO-JP03468.
XX
PR   21-APR-2000; 2000JP-0120358.
XX
PA   (FUSO ) FUSO PHARM IND LTD.
XX
PI   Wakamiya N,  Keshi H,  Ohtani K,  Sakamoto T,  Kishi Y;
XX
DR   WPI; 2002-055345/07.
DR   P-PSDB; ABB56412.
XX
PT   New collectin family proteins, designated CL-L2-1 and CL-L2-2,
PT   expressed in kidney and for treatment and prevention of bacterial and
PT   viral infections -
XX
PS   Claim 20; Page 102-103; 134pp; Japanese.
XX
CC   The invention relates to human collectin family proteins (CL-L2-1 and
CC   CL-L2-2, sequences given in the specification, ABB56407-ABB56411 and
CC   ABB56414-ABB56416), their derivatives and fragments and a related
CC   collectin (CL-L2) of mouse origin (ABB56412) and polynucleotides encoding
CC   all or part of the proteins. The proteins have antibacterial and virucide
CC   activity and are used for protein therapy and treatment, prevention and
CC   diagnosis of bacterial and viral infections. The present sequence is that
CC   of a collectin polynucleotide of the invention.
XX
SQ   Sequence 1522 BP; 425 A; 363 C; 416 G; 318 T; 0 other;


  Query Match             15.1%;  Score 240.6;  DB 24;  Length 1522;
  Best Local Similarity   57.7%;  Pred. No. 3.6e-60;
  Matches  429;  Conservative    0;  Mismatches  314;  Indels    0;  Gaps    0;

Qy    106 ctaccgctgaagtctgtgccacacacacaatttcaccaggacccaaaggagatgatggtg 165
          | || |  || | |||  |    || |   ||   || || | |||||| ||||  || |
Db    239 ccacaggaggacgcctgctctgtgcagattcttgtccccggcctcaaaggggatgcaggag 298

Qy    166 aaaaaggagatccaggagaagagggaaagcatggcaaagtgggacgcatggggccgaaag 225
          |||| |||||   |||||     |||  ||  || | ||| || |  |  ||    ||||
Db    299 aaaagggagacaaaggagccccaggacggccaggaagagtcggccctacaggagaaaaag 358

Qy    226 gaattaaaggagaactgggtgatatgggagatcggggcaatattggcaagactgggccca 285
          ||   |  || ||    ||  | |  ||   |  ||||    ||| || | ||| ||||
Db    359 gagacatgggggacaaaggacagaaaggcactgtgggccgccatggaaaaattggtccca 418

Qy    286 ttgggaagaagggtgacaaagggaaaaaggtttgcttggaatacctggagaaaaaggca 345
          ||||    || ||||| ||||| ||    |||    | |||   |||||     |  ||
Db    419 ttggcgcaaaaggtgaaaaaggagattctggtgatatcggacccctggccccagtggag 478

Qy    346 aagcaggtactgtctgtgattgtggaagataccggaaatttgttggacaactggatatta 405
          || | |||| |   ||||| ||  |        ||||   | ||||  |  |||| |
Db    479 aacctggtattccatgtgagtgcagtcagctgaggaaggctattggggagatggacaacc 538

Qy    406 gtattgcccggctcaagacatctatgaagtttgtcaagaatgtgatagcagggattaggg 465
            | | |  || |  ||     | || ||  | || ||||   | || ||  |  | |
Db    539 aggtcactcaactgacaactgagctaaaattcataaaaaatgctgttgctggcgtgcgcg 598

Qy    466 aaactgaagagaaattctactacatcgtgcaggaagagaagaactacagggaatccctaa 525
          | |||||    ||  |||||    | ||| |||| ||||||   |||   ||  |||
Db    599 agactgagagcaagatctacctgctggtgaaggaggagaagcggtacgcagatgcccagc 658

Qy    526 cccactgcaggattcggggtggaatgctagccatgcccaaggatgaagctgccaacacac 585
              ||||      || || || |  ||   |||||||||| || || || |||||    |
Db    659 tgtcctgccaagcccgaggcggcacactgagcatgcccaaagacgaggcagccaatggcc 718

Qy    586 tcatcgctgactatgttgccaagagtggcttctttcgggtgttcattggcgtgaatgacc 645
          | || |||   ||  | ||  ||  |||| |    || || ||||| ||  | |||||||
Db    719 tgatggcttcatacctggcacaggctggcctggcccgagtcttcatcggtatcaatgacc 778

Qy    646 ttgaaagggagggacagtacatgttcacagacaacactccactgcagaactatagcaact 705
          | || |  || ||    | | ||| | | |||  | | ||  |||||| ||  | ||| |
Db    779 tggagaaagaaggtgctttcgtgtactcggaccgctcccccatgcagaccttcaacaagt 838

Qy    706 ggaatgagggggaacccagcgacccctatggtcatgaggactgtgtggagatgctgagct 765
          ||      || || |||| | || |||||| | | ||||||||||||||||||| ||  ||
Db    839 ggcgcagtggagagcccaacaacgcctatgatgaggaggactgtgtggagatggtggcct 898

Qy    766 ctggcagatggaatgacacagagtgccatcttaccatgtactttgtctgtgagttcatca 825
          | ||  | ||||||||    | |||||  |||||||||||||  | || |||||   ||
Db    899 caggtggctggaatgatgtggcctgccacattaccatgtacttcatgtgcgagtttgaca 958

Qy    826 agaagaaaaagtaacttccctca 848
          |  ||||   || |   ||  ||
Db    959 aagagaacttgtgagagccgaca 981


RESULT  13
ABA91207
ID   ABA91207 standard; DNA; 813 BP.
XX
AC   ABA91207;
XX
DT   19-FEB-2002  (first entry)
XX
DE   Human collectin polynucleotide SEQ ID NO 58.
XX
KW   Human; collectin; CL-L2-1; CL-L2-2; mouse; antibacterial; virucide;
KW   protein therapy; infection; ds.
XX
OS   Homo sapiens.
XX
PN   WO200181401-A1.
XX
PD   01-NOV-2001.
XX
PF   23-APR-2001; 2001WO-JP03468.
XX
PR   21-APR-2000; 2000JP-0120358.
XX
PA   (FUSO ) FUSO PHARM IND LTD.
XX
PI   Wakamiya N,  Keshi H,  Ohtani K,  Sakamoto T,  Kishi Y;
XX
DR   WPI; 2002-055345/07.
XX
PT   New collectin family proteins, designated CL-L2-1 and CL-L2-2,
PT   expressed in kidney and for treatment and prevention of bacterial and
PT   viral infections -
XX
PS   Claim 20; Page 126-127; 134pp; Japanese.
XX
CC   The invention relates to human collectin family proteins (CL-L2-1 and
CC   CL-L2-2, sequences given in the specification, ABB56407-ABB56411 and
CC   ABB56414-ABB56416), their derivatives and fragments and a related
CC   collectin (CL-L2) of mouse origin (ABB56412) and polynucleotides encoding
CC   all or part of the proteins. The proteins have antibacterial and virucide
CC   activity and are used for protein therapy and treatment, prevention and
CC   diagnosis of bacterial and viral infections. The present sequence is that
CC   of a collectin polynucleotide of the invention.
XX
SQ   Sequence 813 BP; 211 A; 198 C; 252 G; 152 T; 0 other;
```

Query Match 15.0%; Score 239; DB 24; Length 813;
Best Local Similarity 58.0%; Pred. No. 7.7e-60;
Matches 422; Conservative 0; Mismatches 305; Indels 0; Gaps 0;

```
Qy    106 ctaccgctgaagtctgtgccacacacacaatttcaccaggacccaaaggagatgatggtg 165
          | || |  || | |||  |    || |   ||   || || | |||||| ||||  || |
Db     83 ccacagaggacgcctgctctgtgcagattcttgtccccggcctcaaaggggatgcaggag 142

Qy    166 aaaaaggagatccaggagaagagggaaagcatggcaaagtgggacgcatggggccgaaag 225
          |||| |||||  |||||     |||  ||  || | ||| || |  |  ||    ||||
Db    143 aaaagggagacaaaggagccccaggacggccaggaagagtcggccctacaggagaaaaag 202

Qy    226 gaattaaaggagaactgggtgatatgggagatcggggcaatattggcaagactgggccca 285
          ||   |  || ||    ||  | |  ||   |  ||||     ||| || | ||| ||||
Db    203 gagacatgggggacaaaggacagaaaggcactgtgggccgccatggaaaaattggtccca 262

Qy    286 ttgggaagaagggtgacaaaggggaaaaaggtttgcttggaatacctggagaaaaaggca 345
          ||||    || ||||| |||||  ||    |||    | |||   |||||    |  ||
Db    263 ttggcgcaaaaggtgaaaaaggagattctggtgatatcggacccccctggccccagtggag 322

Qy    346 aagcaggtactgtctgtgattgtggaagataccggaaatttgttggacaactggatatta 405
          || | |||| |   |||||  ||  |        ||||   | ||||  |  |||| |
Db    323 aacctggtattccatgtgagtgcagtcagctgaggaaggctattggggagatggacaacc 382

Qy    406 gtattgcccggctcaagacatctatgaagtttgtcaagaatgtgatagcagggattaggg 465
          |  | |  || |  ||     | || ||  | || ||||   | || ||  |  | |
Db    383 aggtcactcaactgacaactgagctaaaattcataaaaaatgctgttgctggcgtgcgcg 442

Qy    466 aaactgaagagaaatttctactacatcgtgcaggaagagaagaactacagggaatccctaa 525
          | |||||  . ||  |||||    | ||| |||| ||||||   |||   ||  |||
Db    443 agactgagagcaagatctacctgctggtgaaggaggagaagcggtacgcagatgcccagc 502

Qy    526 cccactgcaggattcggggtggaatgctagccatgcccaaggatgaagctgccaacacac 585
            ||||      || || || |  ||   |||||||||| || || || |||||    |
Db    503 tgtcctgccaagcccgaggcggcacactgagcatgcccaaagacgaggcagccaatggcc 562

Qy    586 tcatcgctgactatgttgccaagagtggcttctttcgggtgttcattggcgtgaatgacc 645
          | || |||  ||  | ||  ||  |||| |    || || ||||| ||  | |||||||
Db    563 tgatggcttcatacctggcacaggctggcctggcccgagtcttcatcggtatcaatgacc 622

Qy    646 ttgaaagggagggacagtacatgttcacagacaacactccactgcagaactatagcaact 705
          | || |  || ||    | | ||| | | |||  | | ||  |||||| ||  | ||| |
Db    623 tggagaaagaaggtgctttcgtgtactcggaccgctccccatgcagaccttcaacaagt 682

Qy    706 ggaatgaggggaacccagcgacccctatggtcatgaggactgtgtggagatgctgagct 765
          ||      || || |||| | || |||||| | | ||||||||||||||||||  ||  ||
Db    683 ggcgcagtggagagcccaacaacgcctatgatgaggaggactgtgtggagatggtggcct 742

Qy    766 ctggcagatggaatgacacagagtgccatcttaccatgtactttgtctgtgagttcatca 825
          | ||  | ||||||||    |  |||||  ||||||||||||  | || |||||   ||
Db    743 caggtggctggaatgatgtggcctgccacattaccatgtacttcatgtgcgagtttgaca 802

Qy    826 agaagaa 832
          |  ||||
Db    803 aagagaa 809
```

RESULT 14
ABA91172
ID ABA91172 standard; cDNA; 1139 BP.
XX
AC ABA91172;
XX
DT 19-FEB-2002 (first entry)
XX
DE Human collectin encoding polynucleotide SEQ ID NO 3.
XX
KW Human; collectin; CL-L2-1; CL-L2-2; mouse; antibacterial; virucide;
KW protein therapy; infection; ss.
XX
OS Homo sapiens.
XX
PN WO200181401-A1.
XX
PD 01-NOV-2001.
XX
PF 23-APR-2001; 2001WO-JP03468.
XX
PR 21-APR-2000; 2000JP-0120358.
XX
PA (FUSO) FUSO PHARM IND LTD.
XX
PI Wakamiya N, Keshi H, Ohtani K, Sakamoto T, Kishi Y;
XX
DR WPI; 2002-055345/07.
DR P-PSDB; ABB56408.
XX
PT New collectin family proteins, designated CL-L2-1 and CL-L2-2,
PT expressed in kidney and for treatment and prevention of bacterial and
PT viral infections -
XX
PS Claim 4; Page 91-92; 134pp; Japanese.
XX
CC The invention relates to human collectin family proteins (CL-L2-1 and
CC CL-L2-2, sequences given in the specification, ABB56407-ABB56411 and
CC ABB56414-ABB56416), their derivatives and fragments and a related
CC collectin (CL-L2) of mouse origin (ABB56412) and polynucleotides encoding
CC all or part of the proteins. The proteins have antibacterial and virucide
CC activity and are used for protein therapy and treatment, prevention and
CC diagnosis of bacterial and viral infections. The present sequence is that
CC of a collectin polynuoleotide of the invention.
XX
SQ Sequence 1139 BP; 251 A; 284 C; 381 G; 223 T; 0 other;

Query Match 14.7%; Score 235.2; DB 24; Length 1139;
Best Local Similarity 58.9%; Pred. No. 1.2e-58;
Matches 405; Conservative 0; Mismatches 283; Indels 0; Gaps 0;

```
Qy    152 aggagatgatggtgaaaaaggagatccaggagaagagggaaagcatggcaaagtgggacg 211
          ||| ||||  || || || |||||   ||| |     |||  || ||| | ||| || |
Db    191 aggggatgcgggagagaagggagacaaaggcgcccccggacggcctggaagagtcggccc 250

Qy    212 catggggccgaaaggaattaaaggagaactgggtgatatgggagatcggggcaatattgg 271
          || |||    ||||||   |  || ||    ||  | |  ||   |  |||    |  |||
Db    251 cacgggagaaaaaggagacatgggggacaaaggacagaaaggcagtgtgggtcgtcatgg 310

Qy    272 caagactgggcccattgggaagaagggtgacaaaggggaaaaaggtttgcttggaatacc 331
          || | ||| ||||||||    || ||||| ||||| ||    |||    | |||   ||
Db    311 aaaaattggtcccattggctctaaaggtgagaaaggagattccggtgacataggaccccc 370

Qy    332 tggagaaaaaggcaaagcaggtactgtctgtgattgtggaagataccggaaatttgttgg 391
          |||    || ||  || ||||       ||||| ||  |       || ||     | ||
Db    371 tggtcctaatggagaaccaggcctcccatgtgagtgcagccagctgcgcaaggccatcgg 430

Qy    392 acaactggatattagtattgcccggctcaagacatctatgaagtttgtcaagaatgtgat 451
            |  |||| |      |  | | ||| |  |      | |||||  |||||||||   |
Db    431 ggagatggacaaccaggtctctcagctgaccagcgagctcaagttcatcaagaatgctgt 490

Qy    452 agcagggattagggaaactgaagagaaattctactacatcgtgcaggaagagaagaacta 511
            || ||  |  | || || ||    || |||||    | ||| |||| ||||||  |||
Db    491 cgccggtgtgcgcgagacggagagcaagatctacctgctggtgaaggaggagaagcgcta 550

Qy    512 cagggaatccctaacccactgcaggattcggggtggaatgctagccatgcccaaggatga 571
          |  |||  |||       ||||  |   || || || | |||   ||||||||||| ||
Db    551 cgcggacgcccagctgtcctgccagggccgcgggggcacgctgagcatgcccaaggacga 610

Qy    572 agctgccaacacactcatcgctgactatgttgccaagagtggcttctttcgggtgttcat 631
          |||||||||    || || || |  ||  | ||  |    ||| |    || || |||||
Db    611 ggctgccaatggcctgatggccgcatacctggcgcaagccggcctggcccgtgtcttcat 670

Qy    632 tggcgtgaatgaccttgaaagggagggacagtacatgttcacagacaacactccactgca 691
          ||| | || |||||  || | ||||||    | | ||| | | ||| || | ||  |||
Db    671 cggcatcaacgacctggagaaggagggcgccttcgtgtactctgaccactcccccatgcg 730
```

```
Qy    692 gaactatagcaactggaatgaggggaacccagcgacccctatggtcatgaggactgtgt 751
          || ||  | ||| |||      || || |||| | |  |||| |   | |||||||| ||
Db    731 gaccttcaacaagtggcgcagcggtgagcccaacaatgcctacgacgaggaggactgcgt 790

Qy    752 ggagatgctgagctctggcagatggaatgacacagagtgccatcttaccatgtactttgt 811
          ||||||| ||  ||| ||| | ||||| |||   |  |||||    |||||||||||  |
Db    791 ggagatggtggcctcgggcggctggaacgacgtggcctgccacaccaccatgtacttcat 850

Qy    812 ctgtgagttcatcaagaagaaaaagtaa 839
            ||||||||   |||| |||| | || |
Db    851 gtgtgagtttgacaaggagaacatgtga 878


RESULT 15
ABA91202
ID   ABA91202 standard; DNA; 735 BP.
XX
AC   ABA91202;
XX
DT   19-FEB-2002  (first entry)
XX
DE   Human collectin polynucleotide SEQ ID NO 46.
XX
KW   Human; collectin; CL-L2-1; CL-L2-2; mouse; antibacterial; virucide;
KW   protein therapy; infection; ds.
XX
OS   Homo sapiens.
XX
PN   WO200181401-A1.
XX
PD   01-NOV-2001.
XX
PF   23-APR-2001; 2001WO-JP03468.
XX
PR   21-APR-2000; 2000JP-0120358.
XX
PA   (FUSO ) FUSO PHARM IND LTD.
XX
PI   Wakamiya N,  Keshi H,  Ohtani K,  Sakamoto T,  Kishi Y;
XX
DR   WPI; 2002-055345/07.
XX
PT   New collectin family proteins, designated CL-L2-1 and CL-L2-2,
PT   expressed in kidney and for treatment and prevention of bacterial and
PT   viral infections
XX
PS   Claim 4; Page 122; 134pp; Japanese.
XX
CC   The invention relates to human collectin family proteins (CL-L2-1 and
CC   CL-L2-2, sequences given in the specification, ABB56407-ABB56411 and
CC   ABB56414-ABB56416), their derivatives and fragments and a related
CC   collectin (CL-L2) of mouse origin (ABB56412) and polynucleotides encoding
CC   all or part of the proteins. The proteins have antibacterial and virucide
CC   activity and are used for protein therapy and treatment, prevention and
CC   diagnosis of bacterial and viral infections. The present sequence is that
CC   of a collectin polynucleotide of the invention.
XX
SQ   Sequence 735 BP; 168 A; 201 C; 249 G; 117 T; 0 other;


  Query Match             14.7%;  Score 233.8;  DB 24;  Length 735;
  Best Local Similarity   58.8%;  Pred. No. 2.5e-58;
  Matches  403;  Conservative    0;  Mismatches  282;  Indels    0;  Gaps    0;

Qy    152 aggagatgatggtgaaaaaggagatccaggagaagagggaaagcatggcaaagtgggacg 211
          ||| ||||  || || || |||||   ||| |     ||| || ||| | ||| || |
Db     51 aggggatgcgggagagaagggagacaaaggcgcccccggacggcctggaagagtcggccc 110

Qy    212 catggggccgaaaggaattaaaggagaactgggtgatatgggagatcggggcaatattgg 271
          || |||    ||||||   |  || ||    ||  | |  ||   |  |||   |  |||
Db    111 cacgggagaaaaaggagacatgggggacaaaggacagaaaggcagtgtgggtcgtcatgg 170

Qy    272 caagactgggcccattgggaagaagggtgacaaaggggaaaaaggtttgcttggaatacc 331
          || | ||| ||||||||    || ||||| ||||| ||    |||    | |||   ||
Db    171 aaaaattggtcccattggctctaaaggtgagaaaggagattccggtgacataggacccccc 230

Qy    332 tggagaaaaaggcaaagcaggtactgtctgtgattgtggaagataccggaaatttgttgg 391
          |||    || ||  || ||||       ||||| || |       || ||     | ||
Db    231 tggtcctaatggagaaccaggcctcccatgtgagtgcagccagctgcgcaaggccatcgg 290

Qy    392 acaactggatattagtattgcccggctcaagacatctatgaagtttgtcaagaatgtgat 451
            |  |||| |      |  | | ||| |  |      | |||||  |||||||||  |
Db    291 ggagatggacaaccaggtctctcagctgaccagcgagctcaagttcatcaagaatgctgt 350

Qy    452 agcagggattagggaaactgaagagaaattctactacatcgtgcaggaagagaagaacta 511
          || ||  |  | || || ||    ||  |||||    | ||| |||| ||||||  |||
Db    351 cgccggtgtgcgcgagacggagagcaagatctacctgctggtgaaggaggagaagcgcta 410

Qy    512 cagggaatccctaacccactgcaggattcggggtggaatgctagccatgcccaaggatga 571
          |  |||  |||       ||||  |   || || || | |||   ||||||||||| ||
Db    411 cgcggacgcccagctgtcctgccagggccgcgggggcacgctgagcatgcccaaggacga 470

Qy    572 agctgccaacacactcatcgctgactatgttgccaagagtggcttctttcgggtgttcat 631
          ||||||||    || || || |  ||  | ||  |    ||| |    || || |||||
Db    471 ggctgccaatggcctgatggccgcatacctggcgcaagccggcctggcccgtgtcttcat 530

Qy    632 tggcgtgaatgaccttgaaagggagggacagtacatgttcacagacaacactccactgca 691
          ||| | || |||||  || | ||||||    | | ||| | | ||| || | ||  |||
Db    531 cggcatcaacgacctggagaaggagggcgccttcgtgtactctgaccactcccccatgcg 590

Qy    692 gaactatagcaactggaatgaggggaacccagcgacccctatggtcatgaggactgtgt 751
          || ||  | ||| |||       || || |||| | |  |||| |   | ||||||| ||
Db    591 gaccttcaacaagtggcgcagcggtgagcccaacaatgcctacgacgaggaggactgcgt 650

Qy    752 ggagatgctgagctctggcagatggaatgacacagagtgccatcttaccatgtactttgt 811
          ||||||| ||  ||| ||| | ||||| |||   |  |||||    ||||||||||  |
Db    651 ggagatggtggcctcgggcggctggaacgacgtggcctgccacaccaccatgtacttcat 710

Qy    812 ctgtgagttcatcaagaagaaaaag 836
            ||||||||   |||| |||| | |
Db    711 gtgtgagtttgacaaggagaacatg 735


Search completed: July  6, 2002, 16:04:13
Job time: 14403 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 6, 2002, 02:20:34 ; Search time 74.83 Seconds
(without alignments)
5235.673 Million cell updates/sec

Title: US-09-600-932-1
Perfect score: 1595
Sequence: 1 cagcaatgaatggctttgca.....gatttaagaaaaacggagcc 1595

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	89.8	5.6	7218	1	US-08-232-463-14	Sequence 14, Appl
	2	68.6	4.3	1868	1	US-08-392-367B-1	Sequence 1, Appli
	3	68.6	4.3	1868	3	US-08-893-467A-1	Sequence 1, Appli
	4	59.6	3.7	1341	2	US-08-945-848-7	Sequence 7, Appli
	5	59.6	3.7	2363	2	US-08-945-848-6	Sequence 6, Appli
	6	57	3.6	1333	4	US-09-227-357-51	Sequence 51, Appl
	7	56.8	3.6	1560	2	US-08-794-795-5	Sequence 5, Appli
	8	56.8	3.6	1560	4	US-09-249-200-5	Sequence 5, Appli
	9	56.8	3.6	1703	2	US-08-794-795-1	Sequence 1, Appli
	10	56.8	3.6	1703	4	US-09-249-200-1	Sequence 1, Appli
	11	56.6	3.5	5102	1	US-08-494-168-1	Sequence 1, Appli
	12	56.4	3.5	1839	1	US-08-383-744-1	Sequence 1, Appli
	13	56.4	3.5	1839	2	US-08-999-336-1	Sequence 1, Appli
	14	56.4	3.5	1839	5	PCT-US96-01427-1	Sequence 1, Appli
	15	51	3.2	1123	3	US-09-188-930-28	Sequence 28, Appl
	16	51	3.2	1123	3	US-09-188-930-203	Sequence 203, App
	17	50.8	3.2	1416	1	US-07-621-091G-1	Sequence 1, Appli
	18	50.8	3.2	1416	2	US-08-399-889-1	Sequence 1, Appli
	19	50.8	3.2	1416	3	US-09-167-364-1	Sequence 1, Appli
	20	50.8	3.2	1416	4	US-09-439-897-1	Sequence 1, Appli
	21	49.4	3.1	885	1	US-08-365-103B-3	Sequence 3, Appli
	22	49.4	3.1	924	1	US-08-365-103B-5	Sequence 5, Appli
	23	49.4	3.1	1005	1	US-08-365-103B-1	Sequence 1, Appli
	24	49.4	3.1	1807	6	5510466-1	Patent No. 5510466
	25	49.2	3.1	1608	4	US-09-029-348-19	Sequence 19, Appl
	26	49	3.1	1881	4	US-09-029-348-20	Sequence 20, Appl
	27	48.6	3.0	1588	6	5510466-3	Patent No. 5510466
	28	48.4	3.0	729	4	US-09-140-804-10	Sequence 10, Appl
	29	48.2	3.0	1276	2	US-08-463-911-1	Sequence 1, Appli
	30	48.2	3.0	4031	1	US-08-159-784-1	Sequence 1, Appli
	31	47.2	3.0	900	4	US-09-198-603C-1	Sequence 1, Appli
	32	46.8	2.9	750	3	US-09-188-930-19	Sequence 19, Appl
	33	46.8	2.9	1107	3	US-09-188-930-217	Sequence 217, App
	34	46.4	2.9	434	4	US-08-991-789A-249	Sequence 249, App
	35	46.4	2.9	434	4	US-09-062-451-249	Sequence 249, App
	36	46.2	2.9	843	4	US-09-118-408-23	Sequence 23, Appl
	37	45.4	2.8	289	4	US-09-007-005-17	Sequence 17, Appl
	38	45.4	2.8	289	4	US-09-244-796-17	Sequence 17, Appl
	39	45.4	2.8	1211	4	US-09-198-603C-25	Sequence 25, Appl
	40	44.8	2.8	3181	1	US-08-655-086-1	Sequence 1, Appli
	41	44.6	2.8	2543	1	US-08-555-669-11	Sequence 11, Appl
	42	44.6	2.8	2543	3	US-09-073-663-11	Sequence 11, Appl
	43	44.4	2.8	7218	1	US-08-232-463-14	Sequence 14, Appl
	44	43.2	2.7	1025	1	US-08-365-103B-9	Sequence 9, Appli
	45	43.2	2.7	1037	1	US-08-365-103B-7	Sequence 7, Appli

ALIGNMENTS

```
RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
;   APPLICANT: DORNER, F.
;   APPLICANT: SCHEIFLINGER, F.
;   APPLICANT: FALKNER, F. G.
;   TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
;   NUMBER OF SEQUENCES: 52
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Foley & Lardner
;     STREET: 1800 Diagonal Road, Suite 500
;     CITY: Alexandria
;     STATE: VA
;     COUNTRY: USA
;     ZIP: 22313-0299
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/232,463
;     FILING DATE:
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US/07/935,313
;     FILING DATE:
;     APPLICATION NUMBER: EP 91 114 300.6
;     FILING DATE: 26-AUG-1991
;   ATTORNEY/AGENT INFORMATION:
;     NAME: BENT, Stephen A.
;     REGISTRATION NUMBER: 29,768
;     REFERENCE/DOCKET NUMBER: 30472/114 IMMU
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (703)836-9300
;     TELEFAX: (703)683-4109
;     TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 7218 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   IMMEDIATE SOURCE:
;     CLONE: pTZgpt-F1s
US-08-232-463-14
```

```
  Query Match             5.6%;  Score 89.8;  DB 1;  Length 7218;
  Best Local Similarity   2.6%;  Pred. No. 1e-17;
  Matches   10;  Conservative  253;  Mismatches  120;  Indels    0;  Gaps    0;

Qy     149 caaaggagatgatggtgaaaaaggagatccaggagaagagggaaagcatggcaaagtggg 208
           |::::::::: :: :: :::::::::::   ::::::::::::::::: : :: :::: :::
Db    1435 CRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1376

Qy     209 acgcatggggccgaaaggaattaaaggagaactgggtgatatgggagatcggggcaatat 268
           : : : ::::  :::::::::  :::::::::  ::: :: : ::::::  :::: :: :
Db    1375 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1316

Qy     269 tggcaagactgggcccattgggaagaagggtgacaaaggggaaaaaggtttgcttggaat 328
           :: ::::  :::   :  :::::::::::: :: ::::::::::::::   :  ::::
Db    1315 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1256

Qy     329 acctggagaaaaaggcaaagcaggtactgtctgtgattgtggaagataccggaaatttgt 388
           :   :::::::::::: :::: ::: :  :   : ::  : :::::: :  :::::   :
Db    1255 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1196

Qy     389 tggacaactggatattagtattgcccggctcaagacatctatgaagtttgtcaagaatgt 448
           ::: ::  ::: :: :: :  :   ::  :::: :   : ::::   :  ::::: :
Db    1195 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1136

Qy     449 gatagcagggattagggaaactgaagagaaattctactacatcgtgcaggaagagaagaa 508
           :: :: ::::  :::::::  :::::::::    :  : :  : : ::::::::::::
Db    1135 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1076

Qy     509 ctacagggaatccctaacccact 531
           : :::::||| | | | | ||
Db    1075 RRRRRRRRRATCGCAAGCTCCCT 1053


RESULT   2
US-08-392-367B-1
; Sequence 1, Application US/08392367B
; Patent No. 5691197
;  GENERAL INFORMATION:
;    APPLICANT:  Tryggvason, Karl
;    APPLICANT:  Elomaa, Outi
;    APPLICANT:  Kangas, Maarit
;    TITLE OF INVENTION:  An Insolated DNA Sequence For a
; Patent No. 5691197
;    TITLE OF INVENTION:  No. 5691197el Macrophage Receptor with
;    TITLE OF INVENTION:  a Collagenous Domain and the
;    TITLE OF INVENTION:  Polypeptide Chain Encoded by
;    TITLE OF INVENTION:  such a Sequence
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fay, Sharpe, Beall, Fagan,
;      ADDRESSEE:  Minnich & McKee
;      STREET:  1100 Superior Avenue
;      STREET:  Suite 700
;      CITY:  Cleveland
;      STATE:  Ohio
;      COUNTRY:  U.S.A.
;      ZIP:  44114-2518
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch,
;      MEDIUM TYPE:  720  Kb storable
;      COMPUTER:  IBM PS/2, Model 35 SX
;      OPERATING SYSTEM:  DOS 5.0
;      SOFTWARE:  Word Perfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/392,367B
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Minnich, Richard J.
;      REGISTRATION NUMBER:  24,175
;      REFERENCE/DOCKET NUMBER:  TRV 2 009
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (216) 861-5582
;      TELEFAX:  (216) 241-1666
;      TELEX:  (216) 980162
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1868 base pairs
;      TYPE:  Nucleic acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
;    MOLECULE TYPE:  Nucleotide-genomic DNA
;    HYPOTHETICAL:  No. 5691197 relevant
;    ANTI-SENSE:  No. 5691197 relevant
US-08-392-367B-1


  Query Match             4.3%;  Score 68.6;  DB 1;  Length 1868;
  Best Local Similarity   56.4%;  Pred. No. 2.1e-11;
  Matches  128;  Conservative    0;  Mismatches   99;  Indels    0;  Gaps    0;

Qy     128 acacacaatttcaccaggacccaaaggagatgatggtgaaaaaggagatccaggagaaga 187
           ||  ||   | |||||||     ||||||    |||||    ||||  || ||||||  |
Db     780 ACTTACGGGTGCACCAGGGAAGCAAGGAGCAACTGGTGCTCCAGGACCTCGAGGAGAGAA 839

Qy     188 gggaaagcatggcaaagtgggacgcatggggccgaaaggaattaaaggagaactgggtga 247
           ||| |   | ||  |  | || | ||  || || || ||      | ||      |||| ||
Db     840 GGGCAGCAAAGGTGACATAGGTCTCACTGGCCCCAAGGGGGAACATGGCACCAAGGGAGA 899

Qy     248 tatgggagatcggggcaatattggcaagactgggcccattgggaagaagggtgacaaagg 307
           |  || || |  |||  |   || || |  |||  ||| || | |||||| ||||  ||
Db     900 CAAAGGGGACCTAGGCCTTCCAGGAAACAAAGGGGACATGGGCATGAAGGGAGACACGGG 959

Qy     308 ggaaaaaggtttgcttggaatacctggagaaaaaggcaaagcaggta 354
           |   |  || |  | ||||   |  ||||  |||||  | || || |
Db     960 GCCCATGGGGTCCCCTGGAGCTCAGGGAGGTAAAGGTGATGCTGGAA 1006


RESULT   3
US-08-893-467A-1
; Sequence 1, Application US/08893467A
; Patent No. 6063901
;  GENERAL INFORMATION:
;    APPLICANT:  Tryggvason, Karl
;    APPLICANT:  Elomaa, Outi
;    APPLICANT:  Kangas, Maarit
;    TITLE OF INVENTION:  An Insolated DNA Sequence For a
; Patent No. 6063901
;    TITLE OF INVENTION:  No. 6063901el Macrophage Receptor with
;    TITLE OF INVENTION:  a Collagenous Domain and the
;    TITLE OF INVENTION:  Polypeptide Chain Encoded by
;    TITLE OF INVENTION:  such a Sequence
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fay, Sharpe, Beall, Fagan,
;      ADDRESSEE:  Minnich & McKee
;      STREET:  1100 Superior Avenue
;      STREET:  Suite 700
;      CITY:  Cleveland
;      STATE:  Ohio
;      COUNTRY:  U.S.A.
;      ZIP:  44114-2518
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch,
;      MEDIUM TYPE:  720  Kb storable
;      COMPUTER:  IBM PS/2, Model 35 SX
;      OPERATING SYSTEM:  DOS 5.0
;      SOFTWARE:  Word Perfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/893,467A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Minnich, Richard J.
;      REGISTRATION NUMBER:  24,175
;      REFERENCE/DOCKET NUMBER:  TRV 2 009
```

```
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (216) 861-5582
;      TELEFAX:  (216) 241-1666
;      TELEX:  (216) 980162
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1868 base pairs
;      TYPE:  Nucleic acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
;    MOLECULE TYPE:  Nucleotide-genomic DNA
;    HYPOTHETICAL:  No. 6063901 relevant
;    ANTI-SENSE:  No. 6063901 relevant
US-08-893-467A-1


  Query Match             4.3%;  Score 68.6;  DB 3;  Length 1868;
  Best Local Similarity   56.4%;  Pred. No. 2.1e-11;
  Matches  128;  Conservative    0;  Mismatches   99;  Indels    0;  Gaps    0;

Qy    128 acacacaatttcaccaggacccaaaggagatgatggtgaaaaaggagatccaggagaaga 187
          ||  ||   | |||||||     ||||||    |||||    ||||  || ||||||  |
Db    780 ACTTACGGGTGCACCAGGGAAGCAAGGAGCAACTGGTGCTCCAGGACCTCGAGGAGAGAA 839

Qy    188 gggaaagcatggcaaagtgggacgcatggggccgaaaggaattaaaggagaaactgggtga 247
          ||| |   | ||  |  | || | ||  || || || ||      | ||       ||| ||
Db    840 GGGCAGCAAAGGTGACATAGGTCTCACTGGCCCCAAGGGGGAACATGGCACCAAGGGAGA 899

Qy    248 tatgggagatcggggcaatattggcaagactgggcccattgggaagaagggtgacaaagg 307
          | || || |  |||  |   || || |  |||  ||| || | |||||| ||||  ||
Db    900 CAAAGGGGACCTAGGCCTTCCAGGAAACAAAGGGGACATGGGCATGAAGGGAGACACGGG 959

Qy    308 ggaaaaaggtttgcttggaatacctggagaaaaaggcaaagcaggta 354
          |   |  || | | ||||   |  ||||  |||||  | || || |
Db    960 GCCCATGGGGTCCCCTGGAGCTCAGGGAGGTAAAGGTGATGCTGGAA 1006


RESULT   4
US-08-945-848-7
; Sequence 7, Application US/08945848
; Patent No. 5968772
;  GENERAL INFORMATION:
;    APPLICANT:  MATSUSHIRO, Aizo
;    TITLE OF INVENTION:  PEARL PROTEIN(NACREIN) AND PROCESS FOR
;    TITLE OF INVENTION:  THE SAME
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  233 South Wacker Drive/6300 Sears Tower
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/945,848
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cawley, Jr, Thomas A.
;      REGISTRATION NUMBER:  40,944
;      REFERENCE/DOCKET NUMBER:  19036/34324
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (312) 474-6300
;      TELEFAX:  (312) 474-0448
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1341 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  21..51
;    FEATURE:
;      NAME/KEY:  binding site
;      LOCATION:  52..99
;    FEATURE:
;      NAME/KEY:  repeat_region
;      LOCATION:  742..969
;    FEATURE:
;      NAME/KEY:  repeat_unit
;      LOCATION:  724..727
;    FEATURE:
;      NAME/KEY:  active site
;      LOCATION:  1123..1266
US-08-945-848-7


  Query Match             3.7%;  Score 59.6;  DB 2;  Length 1341;
  Best Local Similarity   55.2%;  Pred. No. 1.1e-08;
  Matches  116;  Conservative    0;  Mismatches   94;  Indels    0;  Gaps    0;

Qy    144 ggacccaaaggagatgatggtgaaaaaggagatccaggagaagagggaaagcatggcaaa 203
          ||   ||| ||  |  |||||||| || ||  ||     |  |  | ||  |  | || ||
Db    751 GGGGACAACGGTAACAATGGTGACAACGGCAATAACAGCTACAATGGGGACAACGGTAAC 810

Qy    204 gtgggacgcatgggggccgaaaggaattaaaggagaactgggtgatatgggagatcggggc 263
              ||  ||  ||    || ||    || || ||    ||| | |  |||||    |||
Db    811 AATGGTGTCAACGGCAATAACGGCTACAATGGGGACAACGGTAACAATGGAGACAACGGC 870

Qy    264 aatattggcaagactgggcccattgggaagaagggtgacaaaggggaaaaaggtttgctt 323
          ||||  ||| | | ||||  ||  || || || ||||||||| ||  | || |||
Db    871 AATAACGGCTACAATGGGGACAACGGTAACAATGGTGACAACGGCAATAACGGTGAAAAC 930

Qy    324 ggaatacctggagaaaaaggcaaagcaggt 353
          || |     || ||||| |||||    |||
Db    931 GGCAATAACGGTGAAAACGGCAATAACGGT 960


RESULT   5
US-08-945-848-6
; Sequence 6, Application US/08945848
; Patent No. 5968772
;  GENERAL INFORMATION:
;    APPLICANT:  MATSUSHIRO, Aizo
;    TITLE OF INVENTION:  PEARL PROTEIN(NACREIN) AND PROCESS FOR
;    TITLE OF INVENTION:  THE SAME
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  233 South Wacker Drive/6300 Sears Tower
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/945,848
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Cawley, Jr, Thomas A.
;      REGISTRATION NUMBER:  40,944
;      REFERENCE/DOCKET NUMBER:  19036/34324
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (312) 474-6300
```

```
;      TELEFAX:  (312) 474-0448
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2363 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  22..1362
;    FEATURE:
;      NAME/KEY:  sig_peptide
;      LOCATION:  22..72
;    FEATURE:
;      NAME/KEY:  binding site
;      LOCATION:  73..120
;    FEATURE:
;      NAME/KEY:  repeat_region
;      LOCATION:  745..990
;    FEATURE:
;      NAME/KEY:  repeat_unit
;      LOCATION:  745..748
;    FEATURE:
;      NAME/KEY:  active site
;      LOCATION:  1144..1287
US-08-945-848-6


  Query Match             3.7%;  Score 59.6;  DB 2;  Length 2363;
  Best Local Similarity   55.2%;  Pred. No. 1.7e-08;
  Matches  116;  Conservative    0;  Mismatches   94;  Indels    0;  Gaps    0;

Qy    144 ggacccaaaggagatgatggtgaaaaaggagatccaggagaagagggaaagcatggcaaa 203
          ||   ||| ||  |  ||||||| || ||  ||     |  |  | ||  |  | || ||
Db    772 GGGGACAACGGTAACAATGGTGACAACGGCAATAACAGCTACAATGGGGACAACGGTAAC 831

Qy    204 gtgggacgcatggggccgaaaggaattaaaggagaactgggtgatatgggagatcggggc 263
            ||   ||  ||    || ||    || || ||    ||| | |  |||||    |||
Db    832 AATGGTGTCAACGGCAATAACGGCTACAATGGGGACAACGGTAACAATGGAGACAACGGC 891

Qy    264 aatattggcaagactgggcccattgggaagaagggtgacaaaggggaaaaaggtttgctt 323
          ||||  ||| | | ||||  ||  || || || |||||||| ||  | || |||
Db    892 AATAACGGCTACAATGGGGACAACGGTAACAATGGTGACAACGGCAATAACGGTGAAAAC 951

Qy    324 ggaatacctggagaaaaaggcaaagcaggt 353
          || |     || ||||| |||||    |||
Db    952 GGCAATAACGGTGAAAACGGCAATAACGGT 981


RESULT   6
US-09-227-357-51
; Sequence 51, Application US/09227357
; Patent No. 6342581
; GENERAL INFORMATION:
;  APPLICANT: Fischer et al.
;  TITLE OF INVENTION: 123 Human Secreted Proteins
;  FILE REFERENCE: PZ010P1
;  CURRENT APPLICATION NUMBER: US/09/227,357
;  CURRENT FILING DATE: 1999-01-08
;  EARLIER APPLICATION NUMBER: PCT/US98/13684
;  EARLIER FILING DATE: 1998-07-07
;  EARLIER APPLICATION NUMBER: 60/051,926
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/052,793
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/051,925
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/051,929
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/052,803
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/052,732
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/051,931
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/051,932
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/051,916
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/051,930
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/051,918
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/051,920
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/052,733
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/052,795
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/051,919
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/051,928
;  EARLIER FILING DATE: 1997-07-08
;  EARLIER APPLICATION NUMBER: 60/055,722
;  EARLIER FILING DATE: 1997-08-18
;  EARLIER APPLICATION NUMBER: 60/055,723
;  EARLIER FILING DATE: 1997-08-18
;  EARLIER APPLICATION NUMBER: 60/055,948
;  EARLIER FILING DATE: 1997-08-18
;  EARLIER APPLICATION NUMBER: 60/055,949
;  EARLIER FILING DATE: 1997-08-18
;  EARLIER APPLICATION NUMBER: 60/055,953
;  EARLIER FILING DATE: 1997-08-18
;  EARLIER APPLICATION NUMBER: 60/055,950
;  EARLIER FILING DATE: 1997-08-18
;  EARLIER APPLICATION NUMBER: 60/055,947
;  EARLIER FILING DATE: 1997-08-18
;  EARLIER APPLICATION NUMBER: 60/055,964
;  EARLIER FILING DATE: 1997-08-18
;  EARLIER APPLICATION NUMBER: 60/056,360
;  EARLIER FILING DATE: 1997-08-18
;  EARLIER APPLICATION NUMBER: 60/055,684
;  EARLIER FILING DATE: 1997-08-18
;  EARLIER APPLICATION NUMBER: 60/055,984
;  EARLIER FILING DATE: 1997-08-18
;  EARLIER APPLICATION NUMBER: 60/055,954
;  EARLIER FILING DATE: 1997-08-18
;  EARLIER APPLICATION NUMBER: 60/058,785
;  EARLIER FILING DATE: 1997-09-12
;  EARLIER APPLICATION NUMBER: 60/058,664
;  EARLIER FILING DATE: 1997-09-12
;  EARLIER APPLICATION NUMBER: 60/058,660
;  EARLIER FILING DATE: 1997-09-12
;  EARLIER APPLICATION NUMBER: 60/058,661
;  EARLIER FILING DATE: 1997-09-12
;  NUMBER OF SEQ ID NOS: 672
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 51
;  LENGTH: 1333
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: SITE
;  LOCATION: (485)
;  OTHER INFORMATION: n equals a,t,g, or c
;  FEATURE:
;  NAME/KEY: SITE
;  LOCATION: (486)
;  OTHER INFORMATION: n equals a,t,g, or c
;  FEATURE:
;  NAME/KEY: SITE
;  LOCATION: (493)
;  OTHER INFORMATION: n equals a,t,g, or c
;  FEATURE:
;  NAME/KEY: SITE
```

```
;   LOCATION: (496)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (587)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (633)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (1330)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-227-357-51

  Query Match             3.6%;  Score 57;  DB 4;  Length 1333;
  Best Local Similarity   53.5%;  Pred. No. 7.6e-08;
  Matches  114;  Conservative    0;  Mismatches   99;  Indels    0;  Gaps    0;

Qy    152 aggagatgatggtgaaaaaggagatccaggagaagagggaaagcatggcaaagtgggacg 211
          ||||   ||||||  | | ||      ||||||  | ||  |  | || |  |  ||
Db    289 aggaagagatggtagagacggcaggaaaggagagaaaggtgaaaagggaactgcaggttt 348

Qy    212 catggggccgaaaggaattaaaggagaactgggtgatatgggagatcggggcaatattgg 271
          |  ||   ||  |||     |||       ||||| |  || || |  ||  | | |||
Db    349 gagaggtaagactggaccgctaggtcttgccggtgagaaaggggaccaaggagagactgg 408

Qy    272 caagactgggcccattgggaagaagggtgacaaaggggaaaaaggtttgcttggaatacc 331
          ||||  || ||||| ||     |||| || ||||| |||  ||||    ||||    ||
Db    409 gaagaaaggacccataggaccagagggagagaaaggagaagtaggtccaattggtcctcc 468

Qy    332 tggagaaaaaggcaaagcaggtactgtctgtga 364
          ||||  ||| ||  |   | |     | || ||
Db    469 tggaccaaagggagacnnatgatanctntggga 501


RESULT   7
US-08-794-795-5
; Sequence 5, Application US/08794795
; Patent No. 5916766
;  GENERAL INFORMATION:
;    APPLICANT:  Elshourlagy, Nabil
;    APPLICANT:  Adamou, John
;    APPLICANT:  Gross, Mitchell
;    APPLICANT:  Lysko, Paul
;    TITLE OF INVENTION:  Human Macro Scavenger Rec
;    TITLE OF INVENTION:  eptor
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corporation
;      STREET:  709 Swedeland Road
;      CITY:  King of Prussia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19406
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/794,795
;      FILING DATE:  04-FEB-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  ATG50009P
;      FILING DATE:  22-MAY-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Han, William T
;      REGISTRATION NUMBER:  34,344
;      REFERENCE/DOCKET NUMBER:  ATG50009
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-270-5219
;      TELEFAX:  610-270-4026
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1560 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-794-795-5


  Query Match             3.6%;  Score 56.8;  DB 2;  Length 1560;
  Best Local Similarity   54.2%;  Pred. No. 9.7e-08;
  Matches  115;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy    142 caggacccaaaggagatgatggtgaaaaaggagatccaggagaagagggaaagcatggca 201
          |||| |  ||||||    | || || | ||||  || ||||  |  | ||||   | ||
Db   1034 CAGGCCTGAAAGGAAGCAAAGGGGACACAGGACTTCAAGGACAGCAAGGAAGAAAAGGAG 1093

Qy    202 aagtgggacgcatggggccgaaaggaattaaaggagaactgggtgatatgggagatcggg 261
          ||   |||       || ||   |||  | || |||||||  |||        ||  |
Db   1094 AATCAGGAGTTCCAGGCCCTGCAGGTGTGAAGGGAGAACAGGGGAGCCCAGGGCTGGCAG 1153

Qy    262 gcaatattggcaagactgggcccattgggaagaagggtgacaaaggggaaaaaggtttgc 321
          |    |  ||     |||| |    |||  |||||||  ||| | || |  |||| |
Db   1154 GTCCCAAGGGAGCCCCTGGACAAGCTGGCCAGAAGGGAGACCAGGGAGTGAAAGGATCTT 1213

Qy    322 ttggaatacctggagaaaaaggcaaagcaggt 353
          |||    |  |||| ||| ||  ||  ||||
Db   1214 CTGGGGAGCAAGGAGTAAAGGGAGAAAAAGGT 1245


RESULT   8
US-09-249-200-5
; Sequence 5, Application US/09249200
; Patent No. 6197931
;  GENERAL INFORMATION:
;    APPLICANT:  ELSHOURBAGY, NABIL
;    APPLICANT:  ADAMOU, JOHN
;    APPLICANT:  GROSS, MITCHELL
;    APPLICANT:  LYSKO, PAUL
;    TITLE OF INVENTION:  HUMAN MARCO SCAVENGER RECEPTOR
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ratner & Prestia
;      STREET:  P.O. Box 980
;      CITY:  Valley Forge
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19482
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/249,200
;      FILING DATE:  12-FEB-1999
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/794,795
;      FILING DATE:  04-FEB-1997
;      APPLICATION NUMBER:  60/017,699
;      FILING DATE:  23-MAY-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Prestia, Paul F
;      REGISTRATION NUMBER:  23,031
;      REFERENCE/DOCKET NUMBER:  ATG-50009-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-407-0700
```

```
;      TELEFAX:  610-407-0700
;      TELEX:  846169
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1560 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-249-200-5


  Query Match             3.6%;  Score 56.8;  DB 4;  Length 1560;
  Best Local Similarity   54.2%;  Pred. No. 9.7e-08;
  Matches  115;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy     142 caggacccaaaggagatgatggtgaaaaaggagatccaggagaagagggaaagcatggca 201
           |||| |  ||||||    | || || | ||||  || |||| |  | ||||   | ||
Db    1034 CAGGCCTGAAAGGAAGCAAAGGGGACACAGGACTTCAAGGACAGCAAGGAAGAAAAGGAG 1093

Qy     202 aagtgggacgcatggggccgaaaggaattaaaggagaactgggtgatatgggagatcggg 261
           ||   |||      || ||   |||  | || |||||| |||       ||       |
Db    1094 AATCAGGAGTTCCAGGCCCTGCAGGTGTGAAGGGAGAACAGGGGAGCCCAGGGCTGGCAG 1153

Qy     262 gcaatattggcaagactgggcccattgggaagaagggtgacaaaggggaaaaaggtttgc 321
           |    |  ||     |||| |    |||  ||||||| ||| | || |  |||||
Db    1154 GTCCCAAGGGAGCCCCTGGACAAGCTGGCCAGAAGGGAGACCAGGGAGTGAAAGGATCTT 1213

Qy     322 ttggaatacctggagaaaaaggcaaagcaggt 353
           |||    |  |||| ||| ||  ||  ||||
Db    1214 CTGGGGAGCAAGGAGTAAAGGGAGAAAAAGGT 1245


RESULT  9
US-08-794-795-1
; Sequence 1, Application US/08794795
; Patent No. 5916766
;  GENERAL INFORMATION:
;    APPLICANT:  Elshourlagy, Nabil
;    APPLICANT:  Adamou, John
;    APPLICANT:  Gross, Mitchell
;    APPLICANT:  Lysko, Paul
;    TITLE OF INVENTION:  Human Macro Scavenger Rec
;    TITLE OF INVENTION:  eptor
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SmithKline Beecham Corporation
;      STREET:  709 Swedeland Road
;      CITY:  King of Prussia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19406
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/794,795
;      FILING DATE:  04-FEB-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  ATG50009P
;      FILING DATE:  22-MAY-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Han, William T
;      REGISTRATION NUMBER:  34,344
;      REFERENCE/DOCKET NUMBER:  ATG50009
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-270-5219
;      TELEFAX:  610-270-4026
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1703 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-794-795-1


  Query Match             3.6%;  Score 56.8;  DB 2;  Length 1703;
  Best Local Similarity   54.2%;  Pred. No. 1e-07;
  Matches  115;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy     142 caggacccaaaggagatgatggtgaaaaaggagatccaggagaagagggaaagcatggca 201
           |||| |  ||||||    | || || | ||||  || |||| |  | ||||   | ||
Db    1009 CAGGCCTGAAAGGAAGCAAAGGGGACACAGGACTTCAAGGACAGCAAGGAAGAAAAGGAG 1068

Qy     202 aagtgggacgcatggggccgaaaggaattaaaggagaactgggtgatatgggagatcggg 261
           ||   |||      || ||   |||  | || |||||| |||       ||       |
Db    1069 AATCAGGAGTTCCAGGCCCTGCAGGTGTGAAGGGAGAACAGGGGAGCCCAGGGCTGGCAG 1128

Qy     262 gcaatattggcaagactgggcccattgggaagaagggtgacaaaggggaaaaaggtttgc 321
           |    |  ||     |||| |    |||  ||||||| ||| | || |  |||||
Db    1129 GTCCCAAGGGAGCCCCTGGACAAGCTGGCCAGAAGGGAGACCAGGGAGTGAAAGGATCTT 1188

Qy     322 ttggaatacctggagaaaaaggcaaagcaggt 353
           |||    |  |||| ||| ||  ||  ||||
Db    1189 CTGGGGAGCAAGGAGTAAAGGGAGAAAAAGGT 1220


RESULT  10
US-09-249-200-1
; Sequence 1, Application US/09249200
; Patent No. 6197931
;  GENERAL INFORMATION:
;    APPLICANT:  ELSHOURBAGY, NABIL
;    APPLICANT:  ADAMOU, JOHN
;    APPLICANT:  GROSS, MITCHELL
;    APPLICANT:  LYSKO, PAUL
;    TITLE OF INVENTION:  HUMAN MARCO SCAVENGER RECEPTOR
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ratner & Prestia
;      STREET:  P.O. Box 980
;      CITY:  Valley Forge
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19482
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/249,200
;      FILING DATE:  12-FEB-1999
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/794,795
;      FILING DATE:  04-FEB-1997
;      APPLICATION NUMBER:  60/017,699
;      FILING DATE:  23-MAY-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Prestia, Paul F
;      REGISTRATION NUMBER:  23,031
;      REFERENCE/DOCKET NUMBER:  ATG-50009-1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  610-407-0700
;      TELEFAX:  610-407-0700
;      TELEX:  846169
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1703 base pairs
```

```
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-249-200-1


  Query Match             3.6%;  Score 56.8;  DB 4;  Length 1703;
  Best Local Similarity   54.2%;  Pred. No. 1e-07;
  Matches  115;  Conservative    0;  Mismatches   97;  Indels    0;  Gaps    0;

Qy     142 caggacccaaaggagatgatggtgaaaaaggagatccaggagaagagggaaagcatggca 201
           |||| |  ||||||    | || || | ||||  || |||| |  | ||||   | ||
Db    1009 CAGGCCTGAAAGGAAGCAAAGGGGACACAGGACTTCAAGGACAGCAAGGAAGAAAAGGAG 1068

Qy     202 aagtgggacgcatggggccgaaaggaattaaaggagaactgggtgatatgggagatcggg 261
           ||  |||      || ||   |||  | || ||||||| |||       ||       |
Db    1069 AATCAGGAGTTCCAGGCCCTGCAGGTGTGAAGGGAGAACAGGGGAGCCCAGGGCTGGCAG 1128

Qy     262 gcaatattggcaagactgggcccattgggaagaagggtgacaaaggggaaaaaggtttgc 321
           |    |  ||     |||| |    |||  ||||||| ||| | || |  ||||| |
Db    1129 GTCCCAAGGGAGCCCCTGGACAAGCTGGCCAGAAGGGAGACCAGGGAGTGAAAGGATCTT 1188

Qy     322 ttggaatacctggagaaaaaggcaaagcaggt 353
           |||    |  |||| ||| ||  ||  ||||
Db    1189 CTGGGGAGCAAGGAGTAAAGGGAGAAAAAGGT 1220


RESULT  11
US-08-494-168-1
; Sequence 1, Application US/08494168
; Patent No. 5731192
;  GENERAL INFORMATION:
;    APPLICANT:  Reeders, Stephen T.
;    APPLICANT:  Zhou, Jing
;    TITLE OF INVENTION:  Collagen COL4A6: Gene, Protein and Method
;    TITLE OF INVENTION:  of Detecting Collagen Deficiency
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington, D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/494,168
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/112,465
;      FILING DATE:  27-AUG-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SAXE, Bernhard D.
;      REGISTRATION NUMBER:  28,665
;      REFERENCE/DOCKET NUMBER:  40397/104/BABR
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5102 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(2..82, 86..97, 101..4399, 4403..4420, 4424
;      LOCATION:  ..4465, 4469..4876, 4880..5101)
US-08-494-168-1


  Query Match             3.5%;  Score 56.6;  DB 1;  Length 5102;
  Best Local Similarity   54.0%;  Pred. No. 2.5e-07;
  Matches  116;  Conservative    0;  Mismatches   99;  Indels    0;  Gaps    0;

Qy     138 tcaccaggacccaaaggagatgatggtgaaaaaggagatccaggagaagagggaaagcat 197
           | || ||| |||||||| ||    ||       ||    |||||| | | |||| |||
Db    2486 TTACTAGGCCCCAAAGGTGAGCGGGGCAGCCCTGGGACACCAGGACAGGTGGGACAGCCA 2545

Qy     198 ggcaaagtgggacgcatggggccgaaaggaattaaaggagaactgggtgatatgggagat 257
           ||||     |||   |  || ||  | || || || ||  ||   ||       ||||
Db    2546 GGCACCCCAGGATCTAGTGGTCCATATGGCATCAAGGGCAAATCTGGGCTCCCAGGAGCA 2605

Qy     258 cggggcaatattggcaagactgggcccattgggaagaagggtgacaaaggggaaaaaggt 317
           |  |||      ||||   | || |    ||| ||||| ||    | |||  | ||||||
Db    2606 CCAGGCTTCCCAGGCATCTCAGGACATCCTGGAAAGAAAGGAACAAGAGGCAAGAAAGGT 2665

Qy     318 ttgcttggaatacctggagaaaaaggcaaagcagg 352
           | ||||  |  || | | |||||     ||||
Db    2666 CCTCCTGGATCAATTGTAAAGAAAGGGCTGCCAGG 2700


RESULT  12
US-08-383-744-1
; Sequence 1, Application US/08383744
; Patent No. 5702948
;  GENERAL INFORMATION:
;    APPLICANT:  Greene, Mark I.
;    APPLICANT:  Davis, James G.
;    TITLE OF INVENTION:  Saccular collagen and Compositions
;    TITLE OF INVENTION:  and
;    TITLE OF INVENTION:  Methods for Making and Using the Same
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz &
;      ADDRESSEE:  No. 5702948ris
;      STREET:  One LIberty Place, 46th floor
;      CITY:  Philadephia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Wordperfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/383,744
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DeLuca, Mark
;      REGISTRATION NUMBER:  33,229
;      REFERENCE/DOCKET NUMBER:  UPN-2039
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-568-3100
;      TELEFAX:  215-568-3439
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1839 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  both
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  331..1602
US-08-383-744-1
```

```
  Query Match             3.5%;  Score 56.4;  DB 1;  Length 1839;
  Best Local Similarity   53.7%;  Pred. No. 1.4e-07;
  Matches  117;  Conservative    0;  Mismatches  101;  Indels    0;  Gaps    0;

Qy     128 acacacaatttcaccaggacccaaaggagatgatggtgaaaaaggagatccaggagaaga 187
           ||   ||  |  ||| ||| |||| |||    |||| || | ||| ||   ||| ||  |
Db     612 ACCAGCAGGTCTACCTGGAGCCAATGGACTCAATGGCGACATAGGTGAAAAAGGTGATCA 671

Qy     188 gggaaagcatggcaaagtgggacgcatggggccgaaaggaattaaaggagaactgggtga 247
           |||  |  ||       ||  |  |||     |||||    ||||||    |||||
Db     672 AGGACCGGTGGGTCTTCCTGGTGTCCCTGGGATCCCAGGAAAACCAGGAGAGAAAGGTGA 731

Qy     248 tatgggagatcggggcaatattggcaagactgggcccattgggaagaagggtgacaaagg 307
           |   ||       ||  |||  ||  |   |||   || |||   ||| || |||  ||
Db     732 TCCAGGCCTCAAAGGAGATAAAGGTGAACGTGGCTTCAGTGGTCTGAAAGGGGACCCGGG 791

Qy     308 ggaaaaaggtttgcttggaatacctggagaaaaaggca 345
           ||||  |||   || |||  ||  ||||   ||||| |
Db     792 AGAAAGAGGAGAGCCTGGCCTAAATGGAACTAAAGGAA 829


RESULT  13
US-08-999-336-1
; Sequence 1, Application US/08999336
; Patent No. 5891850
;  GENERAL INFORMATION:
;    APPLICANT:  Greene, Mark I.
;    APPLICANT:  Davis, James G.
;    TITLE OF INVENTION:  Saccular collagen and Compositions
;    TITLE OF INVENTION:  and  Methods for Making and Using the Same
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz &
;      ADDRESSEE:  No. 5891850ris
;      STREET:  One LIberty Place, 46th floor
;      CITY:  Philadephia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Wordperfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/999,336
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/383,744
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DeLuca, Mark
;      REGISTRATION NUMBER:  33,229
;      REFERENCE/DOCKET NUMBER:  UPN-2039
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-568-3100
;      TELEFAX:  215-568-3439
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1839 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  both
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  331..1602
US-08-999-336-1


  Query Match             3.5%;  Score 56.4;  DB 2;  Length 1839;
  Best Local Similarity   53.7%;  Pred. No. 1.4e-07;
  Matches  117;  Conservative    0;  Mismatches  101;  Indels    0;  Gaps    0;

Qy     128 acacacaatttcaccaggacccaaaggagatgatggtgaaaaaggagatccaggagaaga 187
           ||   ||  |  ||| ||| |||| |||    |||| || | ||| ||   ||| ||  |
Db     612 ACCAGCAGGTCTACCTGGAGCCAATGGACTCAATGGCGACATAGGTGAAAAAGGTGATCA 671

Qy     188 gggaaagcatggcaaagtgggacgcatggggccgaaaggaattaaaggagaactgggtga 247
           |||  |  ||       ||  |  |||     |||||    ||||||    |||||
Db     672 AGGACCGGTGGGTCTTCCTGGTGTCCCTGGGATCCCAGGAAAACCAGGAGAGAAAGGTGA 731

Qy     248 tatgggagatcggggcaatattggcaagactgggcccattgggaagaagggtgacaaagg 307
           |   ||       ||  |||  ||  |   |||   || |||   ||| || |||  ||
Db     732 TCCAGGCCTCAAAGGAGATAAAGGTGAACGTGGCTTCAGTGGTCTGAAAGGGGACCCGGG 791

Qy     308 ggaaaaaggtttgcttggaatacctggagaaaaaggca 345
           ||||  |||   || |||  ||  ||||   ||||| |
Db     792 AGAAAGAGGAGAGCCTGGCCTAAATGGAACTAAAGGAA 829


RESULT  14
PCT-US96-01427-1
; Sequence 1, Application PC/TUS9601427
;  GENERAL INFORMATION:
;    APPLICANT:  Greene, Mark I.
;    APPLICANT:  Davis, James G.
;    TITLE OF INVENTION:  Saccular collagen and Compositions and
;    TITLE OF INVENTION:  Methods for Making and Using the Same
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz & Norris
;      STREET:  One LIberty Place, 46th floor
;      CITY:  Philadephia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Wordperfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/01427
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/383,744
;      FILING DATE:  02-FEB-1995
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DeLuca, Mark
;      REGISTRATION NUMBER:  33,229
;      REFERENCE/DOCKET NUMBER:  UPN-2653
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-568-3100
;      TELEFAX:  215-568-3439
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1839 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  both
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  331..1602
PCT-US96-01427-1


  Query Match             3.5%;  Score 56.4;  DB 5;  Length 1839;
  Best Local Similarity   53.7%;  Pred. No. 1.4e-07;
  Matches  117;  Conservative    0;  Mismatches  101;  Indels    0;  Gaps    0;

Qy     128 acacacaatttcaccaggacccaaaggagatgatggtgaaaaaggagatccaggagaaga 187
```

```
          ||.  ||  |  ||| ||| |||| |||    |||| || | ||| ||   ||| ||  |
Db    612 ACCAGCAGGTCTACCTGGAGCCAATGGACTCAATGGCGACATAGGTGAAAAAGGTGATCA 671

Qy    188 gggaaagcatggcaaagtgggacgcatggggccgaaaggaattaaaggagaactgggtga 247
          |||  |   ||       ||   |   |||     |||||    ||||||    |||||
Db    672 AGGACCGGTGGGTCTTCCTGGTGTCCCTGGGATCCCAGGAAAACCAGGAGAGAAAGGTGA 731

Qy    248 tatgggagatcggggcaatattggcaagactgggcccattgggaagaagggtgacaaagg 307
          |   ||        ||  |||  ||  |   |||   || |||   ||| || |||   ||
Db    732 TCCAGGCCTCAAAGGAGATAAAGGTGAACGTGGCTTCAGTGGTCTGAAAGGGGACCCGGG 791

Qy    308 ggaaaaaggtttgcttggaatacctggagaaaaaggca 345
          |||| |||   || |||  ||  ||||   ||||| |
Db    792 AGAAAGAGGAGAGCCTGGCCTAAATGGAACTAAAGGAA 829


RESULT  15
US-09-188-930-28
; Sequence 28, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
;  APPLICANT: Watson, James D.
;  APPLICANT: Strachan, Lorna
;  APPLICANT: Sleeman, Matthew
;  APPLICANT: Onrust, Rene
;  APPLICANT: Murison, James Greg
;  TITLE OF INVENTION: Compositions Isolated From Skin Cells
;  TITLE OF INVENTION: and Methods For Their Use
;  FILE REFERENCE: 11000.1011c1
;  CURRENT APPLICATION NUMBER: US/09/188,930A
;  CURRENT FILING DATE: 1998-11-09
;  NUMBER OF SEQ ID NOS: 348
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 28
;   LENGTH: 1123
;   TYPE: DNA
;   ORGANISM: Rat
US-09-188-930-28


  Query Match             3.2%;  Score 51;  DB 3;  Length 1123;
  Best Local Similarity   56.1%;  Pred. No. 5.3e-06;
  Matches   96;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy    139 caccaggacccaaaggagatgatggtgaaaaaggagatccaggagaagagggaaagcatg 198
          | || ||||||  |||   |  |||      |||| | |  ||| |  | |||||  | |
Db    343 cccctggacccccaggtcctcctggcattccaggaaaccatggaaacaatggaaataacg 402

Qy    199 gcaaagtgggacgcatggggccgaaaggaattaaaggagaactgggtgatatgggagatc 258
          |       || | |   ||| | || ||    ||||||||    || ||  ||||   ||
Db    403 gagccactggccacgaaggggccaagggtgagaaaggagacaaaggcgacctggggcctc 462

Qy    259 ggggcaatattggcaagactgggcccattgggaagaagggtgacaaagggg 309
          | ||  |    ||  ||  ||| ||||  ||  |||||||  ||  |||||
Db    463 gaggggaacgggggcagcatggccccaaaggatagaagggatacccagggg 513


Search completed: July  6, 2002, 15:55:03
Job time: 48869 sec
```

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 5, 2002, 20:05:28 ; Search time 1756.54 Seconds
(without alignments)
12255.709 Million cell updates/sec

Title: US-09-600-932-1
Perfect score: 1595
Sequence: 1 cagcaatgaatggctttgca.....gatttaagaaaaacggagcc 1595

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	721.6	45.2	752	10	BM009998	BM009998 603630745
	2	566.8	35.5	609	10	BM010788	BM010788 603629302
	3	455.8	28.6	955	9	BB612129	BB612129 BB612129
	4	414.2	26.0	492	10	BF078010	BF078010 228226 MA
	5	391	24.5	499	10	BI467460	BI467460 389071 MA
c	6	358.2	22.5	496	10	N74624	N74624 za55c02.s1
	7	351	22.0	362	10	R97480	R97480 yq53h02.r1
	8	327	20.5	352	9	AV654961	AV654961 AV654961
	9	302.6	19.0	654	10	BI067078	BI067078 pgf1n.pk0
	10	298.4	18.7	368	10	W00944	W00944 za55c02.r1
	11	293	18.4	357	9	AV653117	AV653117 AV653117
	12	278.6	17.5	354	9	AW435866	AW435866 75149 MAR
	13	274.8	17.2	380	9	BB869893	BB869893 BB869893
c	14	273.4	17.1	449	10	R97432	R97432 yq53h03.s1
	15	265.4	16.6	361	9	BB869996	BB869996 BB869996
	16	249	15.6	451	9	AW355638	AW355638 pft1c.pk0
	17	241.6	15.1	486	10	BM426695	BM426695 pgf2n.pk0
	18	240.6	15.1	1383	11	AK003121	AK003121 Mus muscu
	19	240	15.0	326	10	R29493	R29493 F1-1006D 22
	20	211	13.2	1426	11	BC009951	BC009951 Homo sapi
	21	172.4	10.8	893	10	BF314316	BF314316 601901046
	22	162.4	10.2	590	9	AV690347	AV690347 AV690347
	23	148.2	9.3	723	10	BE382845	BE382845 601297714
	24	144.4	9.1	640	10	BE383325	BE383325 601298236
	25	144.4	9.1	672	10	BF206254	BF206254 601869264
	26	140.2	8.8	564	9	AI353438	AI353438 zeh0500.s
	27	139.2	8.7	823	10	BI198782	BI198782 602759819
	28	136.8	8.6	602	10	BI442205	BI442205 da137c06.
	29	135.2	8.5	702	10	BF311185	BF311185 601898434
	30	133.2	8.4	644	10	BE262656	BE262656 601151465
	31	132.8	8.3	1012	10	BE260904	BE260904 601153812
	32	131.4	8.2	788	10	BF311981	BF311981 601897832
	33	131	8.2	683	10	BE382433	BE382433 601297261
	34	127.2	8.0	715	10	BE313199	BE313199 601149012
	35	126.2	7.9	642	9	AV655586	AV655586 AV655586
	36	125.8	7.9	613	10	BF312666	BF312666 601898164
	37	119.4	7.5	737	10	BE313410	BE313410 601148828
	38	118.8	7.4	654	10	BE312923	BE312923 601146744
	39	118.6	7.4	767	10	BE260355	BE260355 601151448
c	40	112.8	7.1	400	12	AZ322023	AZ322023 1M0042F19
	41	111.6	7.0	542	10	BE313758	BE313758 601149459
	42	110.8	6.9	626	10	BF316496	BF316496 601902094
	43	108.4	6.8	892	10	BF314275	BF314275 601902884
	44	108.2	6.8	538	10	BE312003	BE312003 601145258
c	45	106	6.6	650	12	AZ401277	AZ401277 1M0167I23

ALIGNMENTS

RESULT 1
BM009998
LOCUS BM009998 752 bp mRNA linear EST 30-OCT-2001
DEFINITION 603630745F1 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5444459 5', mRNA sequence.
ACCESSION BM009998
VERSION BM009998.1 GI:16524352
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 752)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1923 row: j column: 12
High quality sequence stop: 752.
FEATURES Location/Qualifiers
source 1. .752
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5444459"
/clone_lib="NIH_MGC_41"
/tissue_type="amelanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California,

Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
BASE COUNT 232 a 135 c 216 g 169 t
ORIGIN

Query Match 45.2%; Score 721.6; DB 10; Length 752;
Best Local Similarity 99.2%; Pred. No. 3.5e-175;
Matches 746; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

```
Qy    150 aaaggagatgatggtgaaaaaggagatccaggagaagagggaaagcatggcaaagtggga 209
          ||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2 AACCGAGATGATGGTGAAAAAGGAGATCCAGGAGAAGAGGGAAAGCATGGCAAAGTGGGA 61

Qy    210 cgcatggggccgaaaggaattaaaggagaactgggtgatatgggagatcggggcaatatt 269
          ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db     62 CGCATGGGGCCGAAAGGAATTAAAGGAGAACTGGGTGATATGGGAGATCAGGGCAATATT 121

Qy    270 ggcaagactgggcccattgggaagaagggtgacaaaggggaaaaaggtttgcttggaata 329
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    122 GGCAAGACTGGGCCCATTGGGAAGAAGGGTGACAAAGGGGAAAAAGGTTTGCTTGGAATA 181

Qy    330 cctggagaaaaaggcaaagcaggtactgtctgtgattgtggaagataccggaaatttgtt 389
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    182 CCTGGAGAAAAAGGCAAAGCAGGTACTGTCTGTGATTGTGGAAGATACCGGAAATTTGTT 241

Qy    390 ggacaactggatattagtattgcccggctcaagacatctatgaagtttgtcaagaatgtg 449
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    242 GGACAACTGGATATTAGTATTGCTCGGCTCAAGACATCTATGAAGTTTGTCAAGAATGTG 301

Qy    450 atagcagggattagggaaactgaagagaaattctactacatcgtgcaggaagagaagaac 509
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    302 ATAGCAGGGATTAGGGAAACTGAAGAGAAATTCTACTACATCGTGCAGGAAGAGAAGAAC 361

Qy    510 tacagggaatccctaacccactgcaggattcggggtggaatgctagccatgcccaaggat 569
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    362 TACAGGGAATCCCTAACCCACTGCAGGATTCGGGGTGGAATGCTAGCCATGCCCAAGGAT 421

Qy    570 gaagctgccaacacactcatcgctgactatgttgccaagagtggcttctttcgggtgttc 629
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    422 GAAGCTGCCAACACACTCATCGCTGACTATGTTGCCAAGAGTGGCTTCTTTCGGGTGTTC 481

Qy    630 attggcgtgaatgaccttgaaagggagggacagtacatgttcacagacaacactccactg 689
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    482 ATTGGCGTGAATGACCTTGAAAGGGAGGGACAGTACATGTTCACAGACAACACTCCACTG 541

Qy    690 cagaactatagcaactggaatgaggggaacccagcgacccctatggtcatgaggactgt 749
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    542 CAGAACTATAGCAACTGGAATGAGGGGGAACCCAGCGACCCCTATGGTCATGAGGACTGT 601

Qy    750 gtggagatgctgagctctggcagatgg-aatgacacagagtgccatcttaccatgtactt 808
          |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db    602 GTGGAGATGCTGAGCTCTGGCAGATGGAAATGACACAGAGTGCCATCTTACCATGTACTT 661

Qy    809 tgtctgtgagttcatcaagaagaaaaagtaacttccctcatcctacgtatttgctatttt 868
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    662 TGTCTGTGAGTTCATCAAGAAGAAAAAGTAACTT-CCTCATCCTACGTATTTGCTATTTT 720

Qy    869 cctgtgaccgtcattacagttattgttatcca 900
          |||||||||||||||||||||||||||||||
Db    721 CCTGTGACCGTCATTACAGTTATTGTTATCCA 752
```

RESULT 2
BM010788
LOCUS BM010788 609 bp mRNA linear EST 30-OCT-2001
DEFINITION 603629302F1 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5434680 5', mRNA sequence.
ACCESSION BM010788
VERSION BM010788.1 GI:16525142
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 609)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1912 row: c column: 01
High quality sequence stop: 608.
FEATURES Location/Qualifiers
source 1..609
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5434680"
/clone_lib="NIH_MGC_41"
/tissue_type="amelanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
BASE COUNT 188 a 121 c 146 g 154 t
ORIGIN

Query Match 35.5%; Score 566.8; DB 10; Length 609;
Best Local Similarity 99.3%; Pred. No. 2.7e-135;
Matches 590; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

```
Qy    367 gtggaagataccggaaatttgttggacaactggatattagtattgcccggctcaagacat 426
          |||||||||||||||||||||||||||||||||||||||||| ||| ||||||||||||
Db      2 GTGGAAGATACCGGAAATTTGTTGGACAACTGGATATTAGTA-TGCTCGGCTCAAGACAT 60

Qy    427 ctatgaagtttgtcaagaatgtgatagcagggattagggaaactgaagagaaattctact 486
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CTATGAAG-TTGTCAAGAATGTGATAGCAGGGATTAGGGAAACTGAAGAGAAATTCTACT 119

Qy    487 acatcgtgcaggaagagaagaactacagggaatccctaacccactgcaggattcggggtg 546
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    120 ACATCGTGCAGGAAGAGAAGAACTACAGGGAATCCCTAACCCACTGCAGGATTCGGGGTG 179

Qy    547 gaatgctagccatgcccaaggatgaagctgccaacacactcatcgctgactatgttgcca 606
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    180 GAATGCTAGCCATGCCCAAGGATGAAGCTGCCAACACACTCATCGCTGACTATGTTGCCA 239

Qy    607 agagtggcttctttcgggtgttcattggcgtgaatgaccttgaaagggagggacagtaca 666
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    240 AGAGTGGCTTCTTTCGGGTGTTCATTGGCGTGAATGACCTTGAAAGGGAGGGACAGTACA 299

Qy    667 tgttcacagacaacactccactgcagaactatagcaactggaatgaggggaacccagcg 726
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    300 TGTTCACAGACAACACTCCACTGCAGAACTATAGCAACTGGAATGAGGGGAACCCAGCG 359

Qy    727 acccctatggtcatgaggactgtgtggagatgctgagctctggcagatggaatgacacag 786
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    360 ACCCCTATGGTCATGAGGACTGTGTGGAGATGCTGAGCTCTGGCAGATGGAATGACACAG 419

Qy    787 agtgccatcttaccatgtactttgtctgtgagttcatcaagaagaaaaagtaacttccct 846
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    420 AGTGCCATCTTACCATGTACTTTGTCTGTGAGTTCATCAAGAAGAAAAAGTAACTTCCCT 479
```

```
Qy    847 catcctacgtatttgctattttcctgtgaccgtcattacagttattgttatccatccttt 906
          ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db    480 CATCCTACGTATTTGCTATTTTCCTGTGACCGTCATTACAGTTATGGTTATCCATCCTTT 539

Qy    907 ttttcctgattgtactacatttgatctgagtcaacatagctagaaaatgctaaa 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    540 TTTTCCTGATTGTACTACATTTGATCTGAGTCAACATAGCTAGAAAATGCTAAA 593


RESULT  3
BB612129
LOCUS       BB612129                955 bp    mRNA    linear   EST 31-AUG-2001
DEFINITION  BB612129 RIKEN full-length enriched, 14 days embryo liver Mus
            musculus cDNA clone 4432404008 5', mRNA sequence.
ACCESSION   BB612129
VERSION     BB612129.1  GI:15394368
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 955)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Yamanaka,I., Kiyosawa,H., Kondo,S., Saito,T., Shinagawa,A., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K., Arakawa,T.,
            Ishii,Y. and Hayashizaki,Y.
             Mapping of 19032 mouse cDNAs on mouse chromosomes. J. Struct.
            Func. Genomics 2 pre, L72-L86 (2001
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
             e mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .955
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="4432404008"
                     /clone_lib="RIKEN full-length enriched, 14 days embryo
                     liver"
                     /sex="mixed"
                     /tissue_type="liver"
                     /dev_stage="14 days embryo"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 10.0 and subtraction to Rot = 100.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
                     3']. cDNA was cloned into the XhoI and BamHI sites.
                     Vector: a modified pBluescript KS(+) after bulk excision
                     from Lambda FLC I"
BASE COUNT      264 a    180 c    261 g    249 t      1 others
ORIGIN


  Query Match           28.6%;  Score 455.8;  DB 9;  Length 955;
  Best Local Similarity  81.7%;  Pred. No. 1.3e-106;
  Matches  539;  Conservative    0;  Mismatches  118;  Indels    3;  Gaps    1;

Qy      2 agcaatgaatggctttgcatccttgcttcgaagaaaccaatttatcctcctggtactatt 61
          ||  |||||||||||||  |  | ||||||||| |||| ||  || ||  || | |||
Db      5 AGTCATGAATGGCTTTAGAGTCCTGCTTCGAAGCAACCTATCCATGCTGTTGCTGCTAGC 64

Qy     62 tcttttgcaaattcagagtctgggtctggatattgatagccgtcctaccgctgaagtctg 121
          ||| |||||  |||||||||||||||||||| ||||||| ||  |  | || |||||||||
Db     65 TCTCTTGCACTTTCAGAGTCTGGGTCTGGATGTTGATAGTCGATCAGCTGCAGAAGTCTG 124

Qy    122 tgccacacacacaatttcaccaggacccaaaggagatgatggtgaaaaaggagatccagg 181
          ||||||||| || |||||||||||||| ||||| |||||||||||||| ||| ||  ||||
Db    125 TGCCACACATACCATTTCACCAGGACCTAAAGGGGATGATGGTGAAAGAGGTGACACAGG 184

Qy    182 agaagagggaaagcatggcaaagtgggacgcatggggccgaaaggaattaaaggagaact 241
          |||||| || ||| |||||||||||||||  ||| || |||||| | |||||||| ||
Db    185 AGAAGAAGGCAAGGATGGCAAAGTGGGACGCCAGGGACCAAAAGGAGTGAAAGGAGAGCT 244

Qy    242 gggtgatatgggagatcggggcaatattggcaagactgggcccattgggaagaagggtga 301
          |||||||||||||  | ||| |||||||||||| |||| || ||||| ||||||||||
Db    245 GGGTGATATGGGAGCCCAGGGTAATATTGGCAAGTCTGGCCCTATTGGCAAGAAGGGTGA 304

Qy    302 caaaggggaaaaaggtttgcttggaatacctggagaaaaaggcaaagcaggtactgtctg 361
          |||||||||||| ||| |||||||||| ||||||||||||||||||||||||||  ||||
Db    305 CAAAGGGGAAAAGGGTCTGCTTGGAATTCCTGGAGAAAAAGGCAAAGCAGGTACCATCTG 364

Qy    362 tgattgtggaagataccggaaatttgttggacaactggatattagtattgcccggctcaa 421
          |||||||| || |||||||| | |||||||||||||||||||| |||| || || ||
Db    365 TGATTGTGGCAGGTACCGGAAAGTGGTTGGACAACTGGATATTAGTGTTGCTCGTCTTAA 424

Qy    422 gacatctatgaagtttgtcaagaatgtgatagcagggattagggaaactgaagagaaatt 481
          |||||| ||||| ||  |||||||||| ||||||||  ||||||||||||||||||||
Db    425 GACATCAATGAAATTCATCAAGAATGTTATAGCAGGGATCCGGGAAACTGAAGAGAAATT 484

Qy    482 ctactacatcgtgcaggaagagaagaactacagggaatccctaacccactgcaggattcg 541
          ||||||||| |||||||| ||||||||||||||||||||| || |||||||||||||| ||
Db    485 CTACTACATTGTGCAGGAGGAGAAGAACTACAGGGAATCTCTGACCCACTGCAGGATCCG 544

Qy    542 gggtggaatgctagccatgcccaaggatgaagctgccaacacactcatcgctgactatgt 601
          || || |||||||||||||| ||  |||||||  |   |||| || || ||||||||||
Db    545 AGGAGGGATGCTAGCCATGCTCAGAGATGAAGTCGTTGACACCCTTATTGCTGACTATGT 604

Qy    602 tgccaagagtggcttctttcgggtgttcattggcgtgaatgaccttgaaagggagggaca 661
          |||   || |   | | || |     | |   | |   || ||||||| |||| ||| ||
Db    605 CGCC---AGAGTGGTTTCTCAGAGTGTTACATGGGGTCATTACCTTGAGAGGGNGGGGCA 661
```

```
RESULT   4
BF078010
LOCUS       BF078010                492 bp    mRNA    linear   EST 18-OCT-2000
DEFINITION  228226 MARC 2PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   BF078010
VERSION     BF078010.1  GI:10871840
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 492)
  AUTHORS   Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
            and Keele,J.W.
  TITLE     Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 49  row: P  column: 11
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .492
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /clone_lib="MARC 2PIG"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from testis, ovary,
                     endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT      149 a     93 c    152 g     98 t
ORIGIN


  Query Match             26.0%;  Score 414.2;  DB 10;  Length 492;
  Best Local Similarity   90.2%;  Pred. No. 5.7e-96;
  Matches  443;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy      1 cagcaatgaatggctttgcatccttgcttcgaagaaaccaatttatcctcctggtactat 60
          ||||||||  ||||| || | |    | ||||||||||||| || ||||||||||  | |
Db      2 CAGCAATGGGTGGCTCTGGAGCTGGACCTCGAAGAAACCAGTTCATCCTCCTGGTGTTCT 61

Qy     61 ttcttttgcaaattcagagtctgggtctggatattgatagccgtcctaccgctgaagtct 120
          |||||||||| |||||||||||||||||| || || || ||||||||||||||||||||
Db     62 TTCTTTTGCAGATTCAGAGTCTGGGTCTGGACATCGACAGTCGTCCTACCGCTGAAGTCT 121

Qy    121 gtgccacacacacaatttcaccaggacccaaaggagatgatggtgaaaaaggagatccag 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db    122 GTGCCACACACACAATTTCACCAGGACCCAAAGGAGATGATGGTGAAAAAGGAGATACAG 181

Qy    181 gagaagagggaaagcatggcaaagtgggacgcatggggccgaaaggaattaaaggagaac 240
          |||| ||||||||||||||||||||||||||||||||||| |||||||||| || ||||
Db    182 GAGAGGAGGGAAAGCATGGCAAAGTGGGACGCATGGGGCCAAAAGGAATTAAGGGTGAAC 241

Qy    241 tgggtgatatgggagatcggggcaatattggcaagactgggcccattgggaagaagggtg 300
          |||||||||| ||||| | |||||| |||||||||||||||||||||| ||||||||||
Db    242 TGGGTGATATAGGAGACCAGGGCAACATTGGCAAGACTGGGCCCATTGGCAAGAAGGGTG 301

Qy    301 acaaaggggaaaaaggtttgcttggaatacctggagaaaaaggcaaagcaggtactgtct 360
          ||||||| ||||| ||||||| ||| || ||||||||||||||||||||||||||||||
Db    302 ACAAAGGAGAAAAGGGTTTGCCTGGGATGCCTGGAGAAAAAGGCAAAGCAGGTACTGTCT 361

Qy    361 gtgattgtggaagataccggaaatttgttggacaactggatattagtattgcccggctca 420
          |||| || |||||||||||||| |||||||||||||||||||| ||| | || || || |
Db    362 GTGACTGCGGAAGATACCGGAAAGTTGTTGGACAACTGGATATCAGTGTGGCTCGCCTGA 421

Qy    421 agacatctatgaagtttgtcaagaatgtgatagcagggattagggaaactgaagagaaat 480
          ||||||| |||||||||||||||||| |||||||||||| ||||||| ||||||||||
Db    422 AGACATCGATGAAGTTTGTCAAGAATGTCATAGCAGGGATTCGGGAAACCGAAGAGAAAT 481

Qy    481 tctactacatc 491
          ||||||||||||
Db    482 TCTACTACATC 492


RESULT   5
BI467460
LOCUS       BI467460                499 bp    mRNA    linear   EST 22-AUG-2001
DEFINITION  389071 MARC 2PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   BI467460
VERSION     BI467460.1  GI:15280338
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 499)
  AUTHORS   Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
            and Keele,J.W.
  TITLE     Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 142  row: E  column: 10
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .499
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /clone_lib="MARC 2PIG"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from testis, ovary,
                     endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT      145 a     96 c    161 g     97 t
ORIGIN


  Query Match             24.5%;  Score 391;  DB 10;  Length 499;
  Best Local Similarity   89.4%;  Pred. No. 5.6e-90;
  Matches  421;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy      1 cagcaatgaatggctttgcatccttgcttcgaagaaaccaatttatcctcctggtactat 60
          ||||||||  |||||||| | |    | ||||||||||||| || ||||||||||  || |
Db     29 CAGCAATGGGTGGCTTTGGAGCTGGACCTCGAAGAAACCAGTTCATCCTCCTGGGGCTCT 88

Qy     61 ttcttttgcaaattcagagtctgggtctggatattgatagccgtcctaccgctgaagtct 120
          |||||||||| |||||||||||||||||| || || || ||| ||||||||||||||||
Db     89 TTCTTTTGCAGATTCAGAGTCTGGGTCTGGACATCGACAGTCGTGCTACCGCTGAAGTCT 148
```

```
Qy    121 gtgccacacacacaatttcaccaggacccaaaggagatgatggtgaaaaaggagatccag 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db    149 GTGCCACACACACAATTTCACCAGGACCCAAAGGAGATGATGGTGAAAAAGGAGATACAG 208

Qy    181 gagaagagggaaagcatggcaaagtgggacgcatggggccgaaaggaattaaaggagaac 240
          |||| |||||||||||||||||||||||||||||||||| |||||||||| || ||||
Db    209 GAGAGGAGGGAAAGCATGGCAAAGTGGGACGCATGGGGCCAAAAGGAATTAAGGGTGAAC 268

Qy    241 tgggtgatatgggagatcggggcaatattggcaagactgggcccattgggaagaagggtg 300
          |||||||||| ||||| | |||||| ||||||||||||||||||||||| |||||||||||
Db    269 TGGGTGATATAGGAGACCAGGGCAACATTGGCAAGACTGGGCCCATTGGCAAGAAGGGTG 328

Qy    301 acaaaggggaaaaaggtttgcttggaatacctggagaaaaaggcaaagcaggtactgtct 360
          ||||||| ||||| || |||| ||| || ||||||||||||||| ||||||||||||||
Db    329 ACAAAGGAGAAAAGGGGTTGCCTGGGATGCCTGGAGAAAAAGGCGAAGCAGGTACTGTCT 388

Qy    361 gtgattgtggaagataccggaaatttgttggacaactggatattagtattgcccggctca 420
          |||| || ||||||||||||||| |||||||||||||||||| ||| | || || || |
Db    389 GTGACTGCGGAAGATACCGGAAAGTTGTTGGACAACTGGATATCAGTGTGGCTCGCCTGA 448

Qy    421 agacatctatgaagtttgtcaagaatgtgatagcagggattagggaaactg 471
          ||||||| |||||||||||||||||| |||||||||||| ||||||| |
Db    449 AGACATCGATGAAGTTTGTCAAGAATGTCATAGCAGGGATTCGGGAAACCG 499
```

RESULT 6
N74624/c
LOCUS N74624 496 bp mRNA linear EST 19-MAR-1996
DEFINITION za55c02.s1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGE:296450 3', mRNA sequence.
ACCESSION N74624
VERSION N74624.1 GI:1231909
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 496)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: m13 -40 forward
High quality sequence stop: 300.
FEATURES Location/Qualifiers
source 1. .496
/organism="Homo sapiens"
/db_xref="GDB:1241371"
/db_xref="taxon:9606"
/clone="IMAGE:296450"
/clone_lib="Soares fetal liver spleen 1NFLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Pac I; Site_2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 149 a 101 c 120 g 122 t 4 others
ORIGIN

Query Match 22.5%; Score 358.2; DB 10; Length 496;
Best Local Similarity 94.0%; Pred. No. 1.6e-81;
Matches 405; Conservative 0; Mismatches 20; Indels 6; Gaps 3;

```
Qy   1170 gtgtctatcccagtggataatttcccag---tttactggtgatgattaggaaggttgttg 1226
          || | |  || || ||||||| || ||      ||  |||||||||||||||||||||||
Db    442 GTCTATCCCCNAGGGGATAAATTTCCCNAGTTTACCTGGTGATGATTAGGAAGGTTGTTG 383

Qy   1227 --atggttaggctaacctgccctggcccaaag-ccagacatgtacaagggctttctgtga 1283
            | |  |||||||||||||||||||||||| |||||||||||||||||||||||||||
Db    382 GATGGTTAGGGCTAACCTGCCCTGGCCCAAAGCCCAGACATGTACAAGGGCTTTCTGTGA 323

Qy   1284 gcaatgataagatctttgaatccaagatgcccagatgttttaccagtcacaccctatggc 1343
          ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db    322 GCAATGATAAGATCTTTGAATCCAAGATGCCCAGATCTTTTACCAGTCACACCCTATGGC 263

Qy   1344 catggctatacttggaagttctccttgttggcacagacatagaaatgctttaaccccaag 1403
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    262 CATGGCTATACTTGGAAGTTCTCCTTGTTGGCACAGACATAGAAATGCTTTAACCCCAAA 203

Qy   1404 cctttatatgggggacttctagctttgtgtcttgtttcagaccatgtggaatgataaata 1463
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    202 CCTTTATATGGGGGACTTCTAGCTTTGTGTCTTGTTTCAGACCATGTGGAATGATAAATA 143

Qy   1464 ctctttttgtgcttctgatctatcgatttcactaacatataccaagtaggtgctttgaac 1523
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    142 CTCTTTTTGTGCTTCTGATCTATCGATTTCACTAACATATACCAAGTAGGTGCTTTGAAC 83

Qy   1524 ccctttctgtaggctcacaccttaatctcaggcccctatatagtcacactttgatttaag 1583
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     82 CCCTTTCTGTAGGCTCACACCTTAATCTCAGGCCCCTATATAGTCACACTTTGATTTAAG 23

Qy   1584 aaaaacggagc 1594
          ||||| |||||
Db     22 AAAAATGGAGC 12
```

RESULT 7
R97480
LOCUS R97480 362 bp mRNA linear EST 11-SEP-1995
DEFINITION yq53h02.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGE:199539 5' similar to SP:A46466 A46466 MANNOSE-BINDING PROTEIN A PRECURSOR - ;, mRNA sequence.
ACCESSION R97480
VERSION R97480.1 GI:983140
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 362)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and Wilson,R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 1136
High quality sequence stops: 337
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

```
            Insert Length: 1136   Std Error: 0.00
            Seq primer: M13RP1
            High quality sequence stop: 337.
FEATURES              Location/Qualifiers
     source           1. .362
                      /organism="Homo sapiens"
                      /db_xref="GDB:3768589"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:199539"
                      /clone_lib="Soares fetal liver spleen 1NFLS"
                      /sex="male"
                      /dev_stage="20 week-post conception fetus"
                      /lab_host="DH10B (ampicillin resistant)"
                      /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                      with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                      1st strand cDNA was primed with a Pac I - oligo(dT) primer
                      [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTT 3'],
                      double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Pac I and cloned into the Pac I
                      and Eco RI sites of the modified pT7T3 vector.  Library
                      went through one round of normalization. Library
                      constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      100 a     77 c     81 g    104 t
ORIGIN


  Query Match             22.0%;  Score 351;  DB 10;  Length 362;
  Best Local Similarity   99.7%;  Pred. No. 1e-79;
  Matches  362;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy    624 gtgttcattggcgtgaatgaccttgaaagggagggacagtacatgttcacagacaacact 683
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GTGTTCATTGGCGTGAATGACCTTGAAAGGGAGGGACAGTACATGTTCACAGACAACACT 60

Qy    684 ccactgcagaactatagcaactggaatgaggggggaacccagcgacccctatggtcatgag 743
          |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db     61 CCACTGCAGAACTATAGCAACTGGAATGAGGGGGAA-CCAGCGACCCCTATGGTCATGAG 119

Qy    744 gactgtgtggagatgctgagctctggcagatggaatgacacagagtgccatcttaccatg 803
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    120 GACTGTGTGGAGATGCTGAGCTCTGGCAGATGGAATGACACAGAGTGCCATCTTACCATG 179

Qy    804 tactttgtctgtgagttcatcaagaagaaaaagtaacttccctcatcctacgtatttgct 863
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    180 TACTTTGTCTGTGAGTTCATCAAGAAGAAAAAGTAACTTCCCTCATCCTACGTATTTGCT 239

Qy    864 attttcctgtgaccgtcattacagttattgttatccatccttttttttcctgattgtacta 923
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    240 ATTTTCCTGTGACCGTCATTACAGTTATTGTTATCCATCCTTTTTTTCCTGATTGTACTA 299

Qy    924 catttgatctgagtcaacatagctagaaaatgctaaactgaggtatggagcctccatcat 983
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    300 CATTTGATCTGAGTCAACATAGCTAGAAAATGCTAAACTGAGGTATGGAGCCTCCATCAT 359

Qy    984 cat 986
          |||
Db    360 CAT 362


RESULT   8
AV654961
LOCUS       AV654961                352 bp    mRNA    linear   EST 15-JAN-2002
DEFINITION  AV654961 GLC Homo sapiens cDNA clone GLCEBC11 3', mRNA sequence.
ACCESSION   AV654961
VERSION     AV654961.1  GI:9875975
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 352)
  AUTHORS   Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
            Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
            Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
            Hu,G., Gu,J., Chen,Z. and Han,Z.
  TITLE     Insight into hepatocellular carcinogenesis at transcriptome level
            by comparing gene expression profiles of hepatocellular carcinoma
            with those of corresponding noncancerous liver
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
  MEDLINE   21625106
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919(ex.45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES              Location/Qualifiers
     source           1. .352
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="GLCEBC11"
                      /clone_lib="GLC"
                      /tissue_type="corresponding non cancerous liver tissue"
                      /dev_stage="Adult"
                      /lab_host="SOLR"
                      /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                      XhoI"
BASE COUNT       90 a     68 c     76 g    118 t
ORIGIN


  Query Match             20.5%;  Score 327;  DB 9;  Length 352;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-73;
  Matches  327;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1004 ttcatattttcacacatggtatgttattgacccaataactcgccaggttacatgggtctt 1063
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     26 TTCATATTTTCACACATGGTATGTTATTGACCCAATAACTCGCCAGGTTACATGGGTCTT 85

Qy   1064 gagagagaattttaattactaattgtgcacgagatagttggttgtctatatgtcaaatga 1123
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     86 GAGAGAGAATTTTAATTACTAATTGTGCACGAGATAGTTGGTTGTCTATATGTCAAATGA 145

Qy   1124 gttgttctcttggtatttgctctaccatctctccctagagcactctgtgtctatcccagt 1183
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    146 GTTGTTCTCTTGGTATTTGCTCTACCATCTCTCCCTAGAGCACTCTGTGTCTATCCCAGT 205

Qy   1184 ggataatttcccagtttactggtgatgattaggaaggttgttgatggttaggctaacctg 1243
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    206 GGATAATTTCCCAGTTTACTGGTGATGATTAGGAAGGTTGTTGATGGTTAGGCTAACCTG 265

Qy   1244 ccctggcccaaagccagacatgtacaagggctttctgtgagcaatgataagatctttgaa 1303
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    266 CCCTGGCCCAAAGCCAGACATGTACAAGGGCTTTCTGTGAGCAATGATAAGATCTTTGAA 325

Qy   1304 tccaagatgcccagatgttttaccagt 1330
          |||||||||||||||||||||||||||
Db    326 TCCAAGATGCCCAGATGTTTTACCAGT 352


RESULT   9
BI067078
LOCUS       BI067078                654 bp    mRNA    linear   EST 15-JUN-2001
DEFINITION  pgf1n.pk010.i8 normalized chicken fat cDNA library Gallus gallus
            cDNA clone pgf1n.pk010.i8 5' similar to gi|5453619 ref|NP_006429.1|
            collectin sub-family member 10 (C-type lectin); collectin liver 1
            [Homo sapiens] dbj|BAA81747.1| (AB002631) collectin 34 [Homo
            sapiens]G, mRNA sequence.
ACCESSION   BI067078
VERSION     BI067078.1  GI:14474600
KEYWORDS    EST.
SOURCE      chicken.
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 654)
  AUTHORS   Cogburn,L.A., Morgan,R.W. and Burnside,J.
  TITLE     Chicken ESTs from fat
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Larry A. Cogburn
            University of Delaware
            Townsend Hall, Newark, DE 19717, USA
            Tel: 302-831-1335
            Fax: 302-831-2822
            Email: cogburn@udel.edu, www.chickest.udel.edu.
FEATURES             Location/Qualifiers
     source          1. .654
                     /organism="Gallus gallus"
                     /db_xref="taxon:9031"
                     /clone="pgf1n.pk010.i8"
                     /clone_lib="normalized chicken fat cDNA library"
                     /sex="Male and Female"
                     /tissue_type="fat"
                     /lab_host="E.coli EMDH10B"
                     /note="Vector: pSPORT1"
BASE COUNT      215 a    107 c    179 g    136 t     17 others
ORIGIN


  Query Match           19.0%;  Score 302.6;  DB 10;  Length 654;
  Best Local Similarity  71.7%;  Pred. No. 3.8e-67;
  Matches  386;  Conservative    0;  Mismatches  152;  Indels    0;  Gaps    0;

Qy      30 cgaagaaaccaatttatcctcctggtactatttcttttgcaaattcagagtctgggtctg 89
           | ||| || |    | ||| | || || ||  | || ||| |||||| | | ||| |
Db     115 CTAAGGAAATATGGGACCCTAGTAGTGCTTTTCATCTTCCAAGTTCAGATTTTTGGTTTT 174

Qy      90 gatattgatagccgtcctaccgctgaagtctgtgccacacacacaatttcaccaggaccc 149
           ||| |||| |  || ||||| | || ||||| | |||||||| |||| ||| |||||
Db     175 GATGTTGACAATCGACCTACAACAGATGTCTGCTCGACACACACTATTTTACCTGGACCA 234

Qy     150 aaaggagatgatggtgaaaaaggagatccaggagaagagggaaagcatggcaaagtggga 209
           ||||| |||||||||||||||||||||  |||||||| ||| || || || ||||| |||
Db     235 AAAGGGGATGATGGTGAAAAAGGAGATAGAGGAGAAGTGGGCAAACAAGGAAAAGTTGGA 294

Qy     210 cgcatggggccgaaaggaattaaaggagaactgggtgatatgggagatcggggcaatatt 269
           |  |  || || ||||||| ||||||    |||| ||| | || || | ||| |   ||
Db     295 CCAAAAGGACCTAAAGGAAACAAAGGAACTGTGGGGGATGTCGGTGACCAGGGAATGCTT 354

Qy     270 ggcaagactgggcccattgggaagaagggtgacaaaggggaaaaaggtttgcttggaata 329
           || || |  || || |||||    || ||||||||||| |  |||||  |    ||  |
Db     355 GGGAAAATCGGTCCGATTGGAGGAAAAGGTGACAAAGGAGCCAAAGGCATATCAGGGGTG 414

Qy     330 cctggagaaaaaggcaaagcaggtactgtctgtgattgtggaagataccggaaatttgtt 389
           |||||| ||||||| |||||||| || |||||||| |||||||| ||||| | | |||||
Db     415 TCTGGAAAAAAAGGAAAAGCAGGCACAGTCTGTGACTGTGGAAGGTACCGCAGAGTTGTT 474

Qy     390 ggacaactggatattagtattgcccggctcaagacatctatgaagtttgtcaagaatgtg 449
           ||||||||| |||| | | |||| ||||| || ||||| || |||||||| ||||||||
Db     475 GGACAACTGAATATCAATGTTGCTCGGCTTAACACATCCATCAAGTTTGTAAAGAATGTT 534

Qy     450 atagcagggattagggaaactgaagagaaattctactacatcgtgcaggaagagaagaac 509
           |||||||| || ||||| || || || |||||||||| || ||  | ||||||||||
Db     535 ATAGCAGGCATCAGGGAGACGGATGAAAAATTCTACTATATTGTCAAAGAAGAGAAGAAT 594

Qy     510 tacagggaatccctaacccactgcaggattcggggtggaatgctagccatgcccaagg 567
           ||||| ||| |||| |  || |||      |      |   || |||||||| || |
Db     595 TACAGAGAAGCCCTGATGCATTGCNNNNNCAGNNNNNNANCACTGGCCATGCCTAANG 652


RESULT  10
W00944
LOCUS       W00944                   368 bp    mRNA    linear   EST 18-APR-1996
DEFINITION  za55c02.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone
            IMAGE:296450 5', mRNA sequence.
ACCESSION   W00944
VERSION     W00944.1  GI:1272942
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 368)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: mob.REGA+ET
            High quality sequence stop: 270.
FEATURES             Location/Qualifiers
     source          1. .368
                     /organism="Homo sapiens"
                     /db_xref="GDB:1241371"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:296450"
                     /clone_lib="Soares fetal liver spleen 1NFLS"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                     1st strand cDNA was primed with a Pac I - oligo(dT) primer
                     [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Pac I and cloned into the Pac I
                     and Eco RI sites of the modified pT7T3 vector.  Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT       94 a     73 c     64 g    136 t      1 others
ORIGIN


  Query Match           18.7%;  Score 298.4;  DB 10;  Length 368;
  Best Local Similarity  97.0%;  Pred. No. 3.8e-66;
  Matches  326;  Conservative    0;  Mismatches    6;  Indels    4;  Gaps    2;

Qy     789 tgccatcttaccatgtactttgtctgtgagttcatcaagaagaaaaagtaacttccctca 848
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 TGCCATCTTACCATGTACTTTGTCTGTGAGTTCATCAAGAAGAAAAAGTAACTTCCCTCA 60

Qy     849 tcctacgtatttgctatttttcctgtgaccgtcattacagttattgttatccatccttttt 908
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TCCTACGTATTTGCTATTTTCCTGTGACCGTCATTACAGTTATTGTTATCCATCCTTTTT 120

Qy     909 ttcctgattgtactacatttgatctgagtcaacatagctagaaaatgctaaactgaggta 968
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TTCCTGATTGTACTACATTTGATCTGAGTCAACATAGCTAGAAAATGCTAAACTGAGGTA 180

Qy     969 tggagcctccatcatcatgctcttttgtgatgattttcatattttcacacatggtatgtt 1028
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db     181 TGGAGCCTCCATCATCATGCTCTTTTGTGATGATTTTCATATTTTCACACATGGTATATT 240

Qy    1029 attgacccaataactcgccaggttacatgggtcttgagagagaattttaatta-ctaatt 1087
           |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db     241 ATTGACCCAATAACTCGCCAGGTTACATGGGTCTTGAGAGAGAATTTTAATTACCTAATT 300

Qy    1088 gtgcac---gagatagttggttgtctatatgtcaaa 1120
           ||||||   || |   | ||||||||||||||||||
Db     301 GTGCACCGAGATAAGTTGGGTTGTCTATATGTCAAA 336
```

```
RESULT 11
AV653117
LOCUS       AV653117                357 bp    mRNA    linear   EST 15-JAN-2002
DEFINITION  AV653117 GLC Homo sapiens cDNA clone GLCDHE09 3', mRNA sequence.
ACCESSION   AV653117
VERSION     AV653117.1  GI:9874131
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 357)
  AUTHORS   Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
            Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
            Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
            Hu,G., Gu,J., Chen,Z. and Han,Z.
  TITLE     Insight into hepatocellular carcinogenesis at transcriptome level
            by comparing gene expression profiles of hepatocellular carcinoma
            with those of corresponding noncancerous liver
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
  MEDLINE   21625106
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919(ex.45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES             Location/Qualifiers
     source          1. .357
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="GLCDHE09"
                     /clone_lib="GLC"
                     /tissue_type="corresponding non cancerous liver tissue"
                     /dev_stage="Adult"
                     /lab_host="SOLR"
                     /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT       88 a     73 c     64 g    132 t
ORIGIN
```

```
  Query Match             18.4%;  Score 293;  DB 9;  Length 357;
  Best Local Similarity   86.9%;  Pred. No. 9.3e-65;
  Matches  357;  Conservative    0;  Mismatches    0;  Indels   54;  Gaps    1;

Qy    765 tctggcagatggaatgacacagagtgccatcttaccatgtactttgtctgtgagttcatc 824
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TCTGGCAGATGGAATGACACAGAGTGCCATCTTACCATGTACTTTGTCTGTGAGTTCATC 60

Qy    825 aagaagaaaaagtaacttccctcatcctacgtatttgctattttcctgtgaccgtcatta 884
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AAGAAGAAAAAGTAACTTCCCTCATCCTACGTATTTGCTATTTTCCTGTGACCGTCATTA 120

Qy    885 cagttattgttatccatcctttttttcctgattgtactacatttgatctgagtcaacata 944
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CAGTTATTGTTATCCATCCTTTTTTTCCTGATTGTACTACATTTGATCTGAGTCAACATA 180

Qy    945 gctagaaaatgctaaactgaggtatggagcctccatcatcatgctcttttgtgatgattt 1004
          |||||                                                      |
Db    181 GCTAG----------------------------------------------------T 186

Qy   1005 tcatattttcacacatggtatgttattgacccaataactcgccaggttacatgggtcttg 1064
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    187 TCATATTTTCACACATGGTATGTTATTGACCCAATAACTCGCCAGGTTACATGGGTCTTG 246

Qy   1065 agagagaattttaattactaattgtgcacgagatagttggttgtctatatgtcaaatgag 1124
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    247 AGAGAGAATTTTAATTACTAATTGTGCACGAGATAGTTGGTTGTCTATATGTCAAATGAG 306

Qy   1125 ttgttctcttggtatttgctctaccatctctccctagagcactctgtgtct 1175
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    307 TTGTTCTCTTGGTATTTGCTCTACCATCTCTCCCTAGAGCACTCTGTGTCT 357
```

```
RESULT 12
AW435866
LOCUS       AW435866                354 bp    mRNA    linear   EST 09-JUL-2000
DEFINITION  75149 MARC 2PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   AW435866
VERSION     AW435866.1  GI:6971244
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 354)
  AUTHORS   Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
            and Keele,J.W.
  TITLE     Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 20
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 30  row: K  column: 4
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .354
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /clone_lib="MARC 2PIG"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from testis, ovary,
                     endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT      101 a     76 c    113 g     64 t
ORIGIN
```

```
  Query Match             17.5%;  Score 278.6;  DB 9;  Length 354;
  Best Local Similarity   89.8%;  Pred. No. 4.8e-61;
  Matches  299;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy      1 cagcaatgaatggctttgcatccttgcttcgaagaaaccaatttatcctcctggtactat 60
          ||||||||  ||||| || | |    | ||||||||||| || |||||||||| || |
Db     22 CAGCAATGGGTGGCTCTGGAGCTGGACCTCGAAGAAACCAGTTCATCCTCCTGGTGCTCT 81

Qy     61 ttcttttgcaaattcagagtctgggtctggatattgatagccgtcctaccgctgaagtct 120
          |||||||||| ||||||||||||||||||| || || || ||||||||||||||||||||
Db     82 TTCTTTTGCAGATTCAGAGTCTGGGTCTGGACATCGACAGTCGTCCTACCGCTGAAGTCT 141

Qy    121 gtgccacacacacaatttcaccaggacccaaaggagatgatggtgaaaaaggagatccag 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db    142 GTGCCACACACACAATTTCACCAGGACCCAAAGGAGATGATGGTGAAAAAGGAGATACAG 201

Qy    181 gagaagagggaaagcatggcaaagtgggacgcatggggccgaaaggaattaaaggagaac 240
          |||| ||||||||||||||||||||||||||||||||||| |||||||||| || ||||
Db    202 GAGAGGAGGGAAAGCATGGCAAAGTGGGACGCATGGGGCCAAAAGGAATTAAGGGTGAAC 261

Qy    241 tgggtgatatgggagatcggggcaatattggcaagactgggcccattgggaagaagggtg 300
```

```
          |||||||||| |||||| | |||||| |||||||||||||||||||||| ||||||||||
Db    262 TGGGTGATATAGGAGACCAGGGCAACATTGGCAAGACTGGGCCCATTGGCAAGAAGGGTG 321

Qy    301 acaaaggggaaaaaggtttgcttggaatacctg 333
          ||||||| ||||| ||||||| ||| || ||||
Db    322 ACAAAGGAGAAAAGGGTTTGCCTGGGATGCCTG 354


RESULT 13
BB869893
LOCUS       BB869893                380 bp    mRNA    linear   EST 27-NOV-2001
DEFINITION  BB869893 RIKEN full-length enriched, pooled tissues, intestinal
            mucosa, etc. Mus musculus cDNA clone G630016H23 5', mRNA sequence.
ACCESSION   BB869893
VERSION     BB869893.1  GI:17116103
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 380)
  AUTHORS   Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
            Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii
            ,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,
            Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.
            , Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
            Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa
            ,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.
            , Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
            2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
             e mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .380
                     /organism="Mus musculus"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="G630016H23"
                     /clone_lib="RIKEN full-length enriched, pooled tissues,
                     intestinal mucosa, etc."
                     /note="pooled tissues ; (tissue_type=intestinal mucosa,
                     dev_stage=adult, sex=male), (tissue_type=accessory
                     axillary lymph node, dev_stage=adult, sex=male),
                     (tissue_type=jejunal and colic lymph node, dev_stage=adult
                     , sex=male), (tissue_type=gall bladder, dev_stage=adult,
                     sex=male), (tissue_type=vesicular gland, dev_stage=adult,
                     sex=male), (tissue_type=spinal cord, dev_stage=11 days
                     embryo), (tissue_type=brain, dev_stage=13 days embryo),
                     (tissue_type=spinal cord, dev_stage=13 days embryo),
                     (tissue_type=lung, dev_stage=14 days embryo),
                     (tissue_type=brain, dev_stage=15 days embryo),
                     (tissue_type=ovary and uterus, dev_stage=10 days pregnant
                     adult, sex=female), (tissue_type=cortex, dev_stage=0 day
                     neonate), (tissue_type=cerebellum, dev_stage=1 month
                     neonate), (tissue_type=diencephalon, dev_stage=16 days
                     neonate, sex=male), (tissue_type=medulla oblongata,
                     dev_stage=16 days neonate, sex=male),
                     (tissue_type=cerebellum, dev_stage=21 days neonate),
                     (tissue_type=testis, dev_stage=8 days neonate, sex=male)"
BASE COUNT      110 a     70 c    123 g     77 t
ORIGIN


  Query Match             17.2%;  Score 274.8;  DB 9;  Length 380;
  Best Local Similarity   83.4%;  Pred. No. 4.7e-60;
  Matches  312;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy      23 cttgcttcgaagaaaccaatttatcctcctggtactatttcttttgcaaattcagagtct 82
           | ||||||||||| |||| ||  || ||  || | |||  ||| |||||  ||||||||||
Db       7 CCTGCTTCGAAGCAACCTATCCATGCTGTTGCTGCTAGCTCTCTTGCACTTTCAGAGTCT 66

Qy      83 gggtctggatattgatagccgtcctaccgctgaagtctgtgccacacacacaatttcacc 142
           |||||||||| ||||||| ||  |  | || ||||||||||||||||| || ||||||||
Db      67 GGGTCTGGATGTTGATAGTCGATCAGCTGCAGAAGTCTGTGCCACACATACCATTTCACC 126

Qy     143 aggacccaaaggagatgatggtgaaaaaggagatccaggagaagagggaaagcatggcaa 202
           |||||| ||||| |||||||||||||| ||| ||  |||||||||| || ||| |||||||
Db     127 AGGACCTAAAGGGGATGATGGTGAAAGAGGTGACACAGGAGAAGAAGGCAAGGATGGCAA 186

Qy     203 agtgggacgcatggggccgaaaggaattaaaggagaactgggtgatatgggagatcgggg 262
           ||||||||||  ||| ||  ||||| | |||||||| ||||||||||||||||    |||
Db     187 AGTGGGACGCCAGGGACCACAAGGAGTGAAAGGAGAGCTGGGTGATATGGGAGCCAAGGG 246

Qy     263 caatattggcaagactgggcccattgggaagaagggtgacaaaggggaaaaaggtttgct 322
           |||||||| ||| |||| || ||||| ||||||||||||||||||||||||| ||| ||||
Db     247 TAATATTGGAAAGTCTGGCCCTATTGGCAAGAAGGGTGACAAAGGGGAAAAGGGTCTGCT 306

Qy     323 tggaatacctggagaaaaaggcaaagcaggtactgtctgtgattgtggaagataccggaa 382
           |||||| |||| ||||||||||||||||||||  |||||||||||| || ||||||||
Db     307 TGGAATTCCTGCAGAAAAAGGCAAAGCAGGTACCATCTGTGATTGTGGCAGGTACCGGAA 366

Qy     383 atttgttggacaac 396
           | | ||||||||||
Db     367 AGTGGTTGGACAAC 380


RESULT 14
R97432/c
LOCUS       R97432                  449 bp    mRNA    linear   EST 11-SEP-1995
DEFINITION  yq53h03.s1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone
            IMAGE:199541 3', mRNA sequence.
ACCESSION   R97432
VERSION     R97432.1  GI:983092
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 449)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
```

```
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 1101
            High quality sequence stops: 301
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1101   Std Error: 0.00
            Seq primer: Promega -21m13
            High quality sequence stop: 301.
FEATURES              Location/Qualifiers
     source           1. .449
                      /organism="Homo sapiens"
                      /db_xref="GDB:3768591"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:199541"
                      /clone_lib="Soares fetal liver spleen 1NFLS"
                      /sex="male"
                      /dev_stage="20 week-post conception fetus"
                      /lab_host="DH10B (ampicillin resistant)"
                      /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                      with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                      1st strand cDNA was primed with a Pac I - oligo(dT) primer
                      [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTT 3'],
                      double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Pac I and cloned into the Pac I
                      and Eco RI sites of the modified pT7T3 vector.  Library
                      went through one round of normalization. Library
                      constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      131 a      92 c      97 g     122 t       7 others
ORIGIN


  Query Match             17.1%;  Score 273.4;  DB 10;  Length 449;
  Best Local Similarity   91.7%;  Pred. No. 1.1e-59;
  Matches  322;  Conservative    0;  Mismatches   23;  Indels    6;  Gaps    3;

Qy   1250 cccaaagccagacatgtacaagggctttctgtgag--caatgataagatctttgaatcca 1307
           |  ||  |||||||||||||||||||||||||||    |  ||||||| ||||||| ||
Db    395 CNAAAGCCCAGACATGTACAAGGGCTTTCTGTGAGGCAATGGATAAGAACTTTGAANCCC 336

Qy   1308 agatgcccaga---tgttttaccagtcaca-ccctatggccatggctatacttggaagtt 1363
           |     ||  |     |||  |||||||| ||||||||||||||||||||||||||||
Db    335 AAGATGCCCCAGATCTTTTACCCAGTCACACCCCTATGGCCATGGCTATACTTGGAAGTT 276

Qy   1364 ctccttgttggcacagacatagaaatgctttaaccccaagcctttatatgggggacttct 1423
           ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db    275 CTCCTTGTTGGCACAGACATAGAAATGCTTTAACCCCAAACCTTTATATGGGGGACTTCT 216

Qy   1424 agctttgtgtcttgtttcagaccatgtggaatgataaatactctttttgtgcttctgatc 1483
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    215 AGCTTTGTGTCTTGTTTCAGACCATGTGGAATGATAAATACTCTTTTTGTGCTTCTGATC 156

Qy   1484 tatcgatttcactaacatataccaagtaggtgctttgaaccccttttctgtaggctcacac 1543
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    155 TATCGATTTCACTAACATATACCAAGTAGGTGCTTTGAACCCCTTTCTGTAGGCTCACAC 96

Qy   1544 cttaatctcaggcccctatatagtcacactttgatttaagaaaaacggagc 1594
           |||||||||||||||||||||||||||||||||||||||||||| |||||
Db     95 CTTAATCTCAGGCCCCTATATAGTCACACTTTGATTTAAGAAAAATGGAGC 45


RESULT  15
BB869996
LOCUS       BB869996                361 bp    mRNA    linear   EST 27-NOV-2001
DEFINITION  BB869996 RIKEN full-length enriched, adult male gall bladder Mus
            musculus cDNA clone G630017H09 5', mRNA sequence.
ACCESSION   BB869996
VERSION     BB869996.1  GI:17116206
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 361)
  AUTHORS   Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
            Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii
            ,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,
            Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.
            , Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
            Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa
            ,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T.
            , Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
            2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
             e mouse tissues.
FEATURES              Location/Qualifiers
     source           1. .361
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="G630017H09"
                      /clone_lib="RIKEN full-length enriched, adult male gall
                      bladder"
                      /sex="male"
                      /tissue_type="gall bladder"
                      /dev_stage="adult"
BASE COUNT      103 a      67 c     117 g      74 t
ORIGIN


  Query Match             16.6%;  Score 265.4;  DB 9;  Length 361;
  Best Local Similarity   84.2%;  Pred. No. 1.2e-57;
  Matches  299;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy     23 cttgcttcgaagaaaccaatttatcctcctggtactatttcttttgcaaattcagagtct 82
          | |||||||||| |||| ||  || ||  || | |||  ||| |||||  ||||||||||
Db      7 CCTGCTTCGAAGCAACCTATCCATGCTGTTGCTGCTAGCTCTCTTGCACTTTCAGAGTCT 66

Qy     83 gggtctggatattgatagccgtcctaccgctgaagtctgtgccacacacacaatttcacc 142
          ||||||||||  ||||||| ||  |  | || ||||||||||||||||| || ||||||||
Db     67 GGGTCTGGATGTTGATAGTCGATCAGCTGCAGAAGTCTGTGCCACACATACCATTTCACC 126

Qy    143 aggacccaaaggagatgatggtgaaaaaggagatccaggagaagagggaaagcatggcaa 202
          ||||||  |||||  |||||||||||||| ||| ||  |||||||||| || ||| |||||||
```

```
Db    127 AGGACCTAAAGGGGATGATGGTGAAAGAGGTGACACAGGAGAAGAAGGCAAGGATGGCAA 186

Qy    203 agtgggacgcatggggccgaaaggaattaaaggagaactgggtgatatgggagatcgggg 262
          ||||||||||  ||| || |||||| | |||||||| ||||||||||||||||  | |||
Db    187 AGTGGGACGCCAGGGACCAAAAGGAGTGAAAGGAGAGCTGGGTGATATGGGAGCCCAGGG 246

Qy    263 caatattggcaagactgggcccattgggaagaagggtgacaaaggggaaaaaggtttgct 322
           |||||||||||| |||| || ||||| ||||||||||||||||||||| ||| ||||
Db    247 TAATATTGGCAAGTCTGGCCCTATTGGCAAGAAGGGTGACAAAGGGGAAAAGGGTCTGCT 306

Qy    323 tggaatacctggagaaaaaggcaaagcaggtactgtctgtgattgtggaagatac 377
          ||||||  ||||||||||||||||||||||||||||  ||||||||||||| || |||
Db    307 TGGAATTCCTGGAGAAAAAGGCAAAGCAGGTACCATCTGTGATTGTGGCAGGTAC 361
```

Search completed: July 6, 2002, 13:21:15
Job time: 62147 sec

GenCore version 4.5

OM nucleic - nucleic search, using sw model

Run on: July 6, 2002, 02:15:59 ; Search time 2948.52 Seconds
(without alignments)
11320.197 Million cell updates/sec

Title: US-09-600-932-1
Perfect score: 1595
Sequence: 1 cagcaatgaatggctttgca.....gatttaagaaaaacggagcc 1595

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1595	100.0	1595	6	E29008	E29008 Novel colle
2	1594	99.9	1594	9	AB002631	AB002631 Homo sapi
3	1141.6	71.6	141262	2	AC080033	AC080033 Homo sapi
c 4	343	21.5	221929	2	AC097055	AC097055 Rattus no
5	246	15.4	1257	9	BC000078	BC000078 Homo sapi
6	158.6	9.9	62064	2	AC107953	AC107953 Homo sapi
7	158.6	9.9	182475	2	AC023487	AC023487 Homo sapi
8	157	9.8	139357	9	AC010907	AC010907 Homo sapi
c 9	157	9.8	204511	2	AC108488	AC108488 Homo sapi
10	110.8	6.9	105156	2	AC106609	AC106609 Rattus no
c 11	95.4	6.0	767	11	HSU92547	U92547 Homo sapien
c 12	89.8	5.6	7218	6	I66494	I66494 Sequence 14
c 13	79	5.0	182475	2	AC023487	AC023487 Homo sapi
14	74	4.6	5851	3	AF282902	AF282902 Hydra vul
15	73.6	4.6	1529	1	SPY301812	AJ301812 Streptoco
16	70.2	4.4	1630	10	AF125191	AF125191 Mesocrice
17	69.6	4.4	199982	2	AL626774	AL626774 Mus muscu
18	68.8	4.3	5174	9	HSCOL4A3	X80031 Homo sapien
19	68.6	4.3	1836	10	MMU18424	U18424 Mus musculu
20	68.6	4.3	1868	6	I76405	I76405 Sequence 1
21	66.6	4.2	804	1	AF296338	AF296338 Streptoco
22	64.6	4.1	4584	5	CHKC6A1A	J04598 Chicken alp
23	64.6	4.1	4595	5	GGCOL6A1	X64458 G.gallus Co
24	63.8	4.0	909	6	AX054812	AX054812 Sequence
25	63.8	4.0	10057	5	CHKCOLAVI	M24282 Chicken alp
26	63.6	4.0	1192	1	AF296337	AF296337 Streptoco
27	63.6	4.0	7765	10	AF169387	AF169387 Mus muscu
28	63.4	4.0	8207	5	AB045975	AB045975 Chrysophr
c 29	63.4	4.0	39183	3	CEM199	Z81104 Caenorhabdi
30	63.2	4.0	2766	10	MUSCOLAIIX	L12215 Mouse colla
31	63.2	4.0	2883	9	HUMCOL11A2	J04974 Human alpha
32	63.2	4.0	3049	10	MUSNA411	D17511 Mouse mRNA
33	63	3.9	6143	3	SPCOLP4	X76730 S.purpuratu
34	62.8	3.9	2160	5	CHKCOLAI2	J04425 Chicken typ
35	62.8	3.9	4153	5	GDCOL6A2	X15041 Chicken mRN
c 36	62.8	3.9	203606	2	AL645990	AL645990 Mus muscu
37	62.4	3.9	6158	6	AX329923	AX329923 Sequence
38	62.4	3.9	6158	6	AX333280	AX333280 Sequence
39	62.4	3.9	6158	9	HUMCA1XIA	J04177 Human alpha
c 40	62.4	3.9	102653	14	CYSCG	L63545 Lymphocysti
c 41	62.2	3.9	105156	2	AC106609	AC106609 Rattus no
42	62	3.9	879	1	AF336808	AF336808 Streptoco
43	62	3.9	9320	1	AE005298	AE005298 Escherich
44	62	3.9	60942	7	AP000363	AP000363 Bacteriop
45	62	3.9	61670	7	AF125520	AF125520 Bacteriop

ALIGNMENTS

RESULT 1
E29008
LOCUS E29008 1595 bp DNA linear PAT 07-FEB-2001
DEFINITION Novel collectin.
ACCESSION E29008
VERSION E29008.1 GI:13018416
KEYWORDS JP 1999206377-A/1.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1595)
AUTHORS Nobutaka,W.
TITLE Novel collectin
JOURNAL Patent: JP 1999206377-A 1 03-AUG-1999;
FUSO YAKUHIN KOGYO KK
COMMENT OS Homo sapiens (human)
PN JP 1999206377-A/1
PD 03-AUG-1999
PF 23-JAN-1998 JP 1998011281
PR
PI NOBUTAKA WAKAMIYA
PC C12N15/09,C07K14/47,C07K14/78,C12P21/00,C12N15/00 CC

```
          Strandedness: Double;
          CC   Topology: Linear;
          FH   Key             Location/Qualifiers
          FT   CDS             6. .836.
FEATURES             Location/Qualifiers
     source          1. .1595
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      444 a    322 c    382 g    447 t
ORIGIN
```

Query Match 100.0%; Score 1595; DB 6; Length 1595;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1595; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 cagcaatgaatggctttgcatccttgcttcgaagaaaccaatttatcctcctggtactat 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CAGCAATGAATGGCTTTGCATCCTTGCTTCGAAGAAACCAATTTATCCTCCTGGTACTAT 60

Qy      61 ttcttttgcaaattcagagtctgggtctggatattgatagccgtcctaccgctgaagtct 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TTCTTTTGCAAATTCAGAGTCTGGGTCTGGATATTGATAGCCGTCCTACCGCTGAAGTCT 120

Qy     121 gtgccacacacacaatttcaccaggacccaaaggagatgatggtgaaaaaggagatccag 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GTGCCACACACACAATTTCACCAGGACCCAAAGGAGATGATGGTGAAAAAGGAGATCCAG 180

Qy     181 gagaagagggaaagcatggcaaagtgggacgcatggggccgaaaggaattaaaggagaaac 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GAGAAGAGGGAAAGCATGGCAAAGTGGGACGCATGGGGCCGAAAGGAATTAAAGGAGAAC 240

Qy     241 tgggtgatatgggagatcggggcaatattggcaagactgggcccattgggaagaagggtg 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TGGGTGATATGGGAGATCGGGGCAATATTGGCAAGACTGGGCCCATTGGGAAGAAGGGTG 300

Qy     301 acaaagggaaaaaggtttgcttggaatacctggagaaaaaggcaaagcaggtactgtct 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 ACAAAGGGAAAAAGGTTTGCTTGGAATACCTGGAGAAAAAGGCAAAGCAGGTACTGTCT 360

Qy     361 gtgattgtggaagataccggaaatttgttggacaactggatattagtattgcccggctca 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GTGATTGTGGAAGATACCGGAAATTTGTTGGACAACTGGATATTAGTATTGCCCGGCTCA 420

Qy     421 agacatctatgaagtttgtcaagaatgtgatagcagggattagggaaactgaagagaaat 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 AGACATCTATGAAGTTTGTCAAGAATGTGATAGCAGGGATTAGGGAAACTGAAGAGAAAT 480

Qy     481 tctactacatcgtgcaggaagagaagaactacagggaatccctaacccactgcaggattc 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 TCTACTACATCGTGCAGGAAGAGAAGAACTACAGGGAATCCCTAACCCACTGCAGGATTC 540

Qy     541 ggggtggaatgctagccatgcccaaggatgaagctgccaacacactcatcgctgactatg 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 GGGGTGGAATGCTAGCCATGCCCAAGGATGAAGCTGCCAACACACTCATCGCTGACTATG 600

Qy     601 ttgccaagagtggcttctttcgggtgttcattggcgtgaatgaccttgaaagggagggac 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 TTGCCAAGAGTGGCTTCTTTCGGGTGTTCATTGGCGTGAATGACCTTGAAAGGGAGGGAC 660

Qy     661 agtacatgttcacagacaacactccactgcagaactatagcaactggaatgagggggaac 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 AGTACATGTTCACAGACAACACTCCACTGCAGAACTATAGCAACTGGAATGAGGGGGAAC 720

Qy     721 ccagcgacccctatggtcatgaggactgtgtggagatgctgagctctggcagatggaatg 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 CCAGCGACCCCTATGGTCATGAGGACTGTGTGGAGATGCTGAGCTCTGGCAGATGGAATG 780

Qy     781 acacagagtgccatcttaccatgtactttgtctgtgagttcatcaagaagaaaaagtaac 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 ACACAGAGTGCCATCTTACCATGTACTTTGTCTGTGAGTTCATCAAGAAGAAAAAGTAAC 840

Qy     841 ttccctcatcctacgtatttgctattttcctgtgaccgtcattacagttattgttatcca 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 TTCCCTCATCCTACGTATTTGCTATTTTCCTGTGACCGTCATTACAGTTATTGTTATCCA 900

Qy     901 tccttttttcctgattgtactacatttgatctgagtcaacatagctagaaaatgctaaa 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 TCCTTTTTTCCTGATTGTACTACATTTGATCTGAGTCAACATAGCTAGAAAATGCTAAA 960

Qy     961 ctgaggtatggagcctccatcatcatgctcttttgtgatgattttcatattttcacacat 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 CTGAGGTATGGAGCCTCCATCATCATGCTCTTTTGTGATGATTTTCATATTTTCACACAT 1020

Qy    1021 ggtatgttattgacccaataactcgccaggttacatgggtcttgagagagaattttaatt 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 GGTATGTTATTGACCCAATAACTCGCCAGGTTACATGGGTCTTGAGAGAGAATTTTAATT 1080

Qy    1081 actaattgtgcacgagatagttggttgtctatatgtcaaatgagttgttctcttggtatt 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 ACTAATTGTGCACGAGATAGTTGGTTGTCTATATGTCAAATGAGTTGTTCTCTTGGTATT 1140

Qy    1141 tgctctaccatctctccctagagcactctgtgtctatcccagtggataatttcccagttt 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 TGCTCTACCATCTCTCCCTAGAGCACTCTGTGTCTATCCCAGTGGATAATTTCCCAGTTT 1200

Qy    1201 actggtgatgattaggaaggttgttgatggttaggctaacctgccctggcccaaagccag 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 ACTGGTGATGATTAGGAAGGTTGTTGATGGTTAGGCTAACCTGCCCTGGCCCAAAGCCAG 1260

Qy    1261 acatgtacaagggctttctgtgagcaatgataagatctttgaatccaagatgcccagatg 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 ACATGTACAAGGGCTTTCTGTGAGCAATGATAAGATCTTTGAATCCAAGATGCCCAGATG 1320

Qy    1321 ttttaccagtcacaccctatggccatggctatacttggaagttctccttgttggcacaga 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 TTTTACCAGTCACACCCTATGGCCATGGCTATACTTGGAAGTTCTCCTTGTTGGCACAGA 1380

Qy    1381 catagaaatgctttaaccccaagcctttatatgggggacttctagctttgtgtcttgttt 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 CATAGAAATGCTTTAACCCCAAGCCTTTATATGGGGGACTTCTAGCTTTGTGTCTTGTTT 1440

Qy    1441 cagaccatgtggaatgataaatactctttttgtgcttctgatctatcgatttcactaaca 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 CAGACCATGTGGAATGATAAATACTCTTTTTGTGCTTCTGATCTATCGATTTCACTAACA 1500

Qy    1501 tataccaagtaggtgctttgaaccccttttctgtaggctcacaccttaatctcaggcccct 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 TATACCAAGTAGGTGCTTTGAACCCCTTTCTGTAGGCTCACACCTTAATCTCAGGCCCCT 1560

Qy    1561 atatagtcacactttgatttaagaaaaacggagcc 1595
          ||||||||||||||||||||||||||||||||||
Db    1561 ATATAGTCACACTTTGATTTAAGAAAAACGGAGCC 1595
```

```
RESULT   2
AB002631
LOCUS       AB002631               1594 bp    mRNA    linear   PRI 24-JUN-1999
DEFINITION  Homo sapiens mRNA for collectin 34, complete cds.
ACCESSION   AB002631
VERSION     AB002631.1  GI:5162874
KEYWORDS    collectin 34.
SOURCE      Homo sapiens cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Ohtani,K., Suzuki,Y., Eda,S., Kawai,T., Kase,T., Yamazaki,H.,
            Keshi,H., Sakai,Y., Fukuoh,A., Sakamoto,T. and Wakamiya,N.
  TITLE     Molecular cloning of a novel human collectin from liver (CL-L1)
  JOURNAL   J. Biol. Chem. 274 (19), 13681-13689 (1999)
  MEDLINE   99240768
REFERENCE   2  (bases 1 to 1594)
  AUTHORS   Ohtani,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-APR-1997) Katsuki Ohtani, Osaka Prefectural Institute
```

of Public Health, Department of Pathology; 3-69,Nakamichi 1-chome
Higashinari-ku, Osaka, Osaka 537, Japan
(E-mail:suzuki@iph.pref.osaka.jp, Tel:+81-6-972-1321,
Fax:+81-6-972-0772)

```
FEATURES             Location/Qualifiers
     source          1. .1594
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             6. .839
                     /codon_start=1
                     /product="collectin 34"
                     /protein_id="BAA81747.1"
                     /db_xref="GI:5162875"
                     /translation="MNGFASLLRRNQFILLVLFLLQIQSLGLDIDSRPTAEVCATHTI
                     SPGPKGDDGEKGDPGEEGKHGKVGRMGPKGIKGELGDMGDRGNIGKTGPIGKKGDKGE
                     KGLLGIPGEKGKAGTVCDCGRYRKFVGQLDISIARLKTSMKFVKNVIAGIRETEEKFY
                     YIVQEEKNYRESLTHCRIRGGMLAMPKDEAANTLIADYVAKSGFFRVFIGVNDLEREG
                     QYMFTDNTPLQNYSNWNEGEPSDPYGHEDCVEMLSSGRWNDTECHLTMYFVCEFIKKK
                     K"
     polyA_signal    1037. .1041
BASE COUNT      444 a    321 c    382 g    447 t
ORIGIN
```

Query Match 99.9%; Score 1594; DB 9; Length 1594;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1594; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 cagcaatgaatggctttgcatccttgcttcgaagaaaccaatttatcctcctggtactat 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CAGCAATGAATGGCTTTGCATCCTTGCTTCGAAGAAACCAATTTATCCTCCTGGTACTAT 60

Qy      61 ttcttttgcaaattcagagtctgggtctggatattgatagccgtcctaccgctgaagtct 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TTCTTTTGCAAATTCAGAGTCTGGGTCTGGATATTGATAGCCGTCCTACCGCTGAAGTCT 120

Qy     121 gtgccacacacacaatttcaccaggacccaaaggagatgatggtgaaaaaggagatccag 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GTGCCACACACACAATTTCACCAGGACCCAAAGGAGATGATGGTGAAAAAGGAGATCCAG 180

Qy     181 gagaagagggaaagcatggcaaagtgggacgcatggggccgaaaggaattaaaggagaac 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GAGAAGAGGGAAAGCATGGCAAAGTGGGACGCATGGGGCCGAAAGGAATTAAAGGAGAAC 240

Qy     241 tgggtgatatgggagatcggggcaatattggcaagactgggcccattgggaagaagggtg 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TGGGTGATATGGGAGATCGGGGCAATATTGGCAAGACTGGGCCCATTGGGAAGAAGGGTG 300

Qy     301 acaaaggggaaaaaggtttgcttggaatacctggagaaaaaggcaaagcaggtactgtct 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 ACAAAGGGGAAAAAGGTTTGCTTGGAATACCTGGAGAAAAAGGCAAAGCAGGTACTGTCT 360

Qy     361 gtgattgtggaagataccggaaatttgttggacaactggatattagtattgcccggctca 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GTGATTGTGGAAGATACCGGAAATTTGTTGGACAACTGGATATTAGTATTGCCCGGCTCA 420

Qy     421 agacatctatgaagtttgtcaagaatgtgatagcagggattagggaaactgaagagaaat 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 AGACATCTATGAAGTTTGTCAAGAATGTGATAGCAGGGATTAGGGAAACTGAAGAGAAAT 480

Qy     481 tctactacatcgtgcaggaagagaagaactacagggaatccctaacccactgcaggattc 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 TCTACTACATCGTGCAGGAAGAGAAGAACTACAGGGAATCCCTAACCCACTGCAGGATTC 540

Qy     541 ggggtggaatgctagccatgcccaaggatgaagctgccaacacactcatcgctgactatg 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 GGGGTGGAATGCTAGCCATGCCCAAGGATGAAGCTGCCAACACACTCATCGCTGACTATG 600

Qy     601 ttgccaagagtggcttctttcgggtgttcattggcgtgaatgaccttgaaagggagggac 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 TTGCCAAGAGTGGCTTCTTTCGGGTGTTCATTGGCGTGAATGACCTTGAAAGGGAGGGAC 660

Qy     661 agtacatgttcacagacaacactccactgcagaactatagcaactggaatgaggggaac 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 AGTACATGTTCACAGACAACACTCCACTGCAGAACTATAGCAACTGGAATGAGGGGAAC 720

Qy     721 ccagcgacccctatggtcatgaggactgtgtggagatgctgagctctggcagatggaatg 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 CCAGCGACCCCTATGGTCATGAGGACTGTGTGGAGATGCTGAGCTCTGGCAGATGGAATG 780

Qy     781 acacagagtgccatcttaccatgtactttgtctgtgagttcatcaagaagaaaaagtaac 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 ACACAGAGTGCCATCTTACCATGTACTTTGTCTGTGAGTTCATCAAGAAGAAAAAGTAAC 840

Qy     841 ttccctcatcctacgtatttgctattttcctgtgaccgtcattacagttattgttatcca 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 TTCCCTCATCCTACGTATTTGCTATTTTCCTGTGACCGTCATTACAGTTATTGTTATCCA 900

Qy     901 tccttttttttcctgattgtactacatttgatctgagtcaacatagctagaaaaatgctaaa 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 TCCTTTTTTTCCTGATTGTACTACATTTGATCTGAGTCAACATAGCTAGAAAATGCTAAA 960

Qy     961 ctgaggtatggagcctccatcatcatgctcttttgtgatgattttcatattttcacacat 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 CTGAGGTATGGAGCCTCCATCATCATGCTCTTTTGTGATGATTTTCATATTTTCACACAT 1020

Qy    1021 ggtatgttattgacccaataactcgccaggttacatgggtcttgagagagaatttttaatt 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 GGTATGTTATTGACCCAATAACTCGCCAGGTTACATGGGTCTTGAGAGAGAATTTTAATT 1080

Qy    1081 actaattgtgcacgagatagttggttgtctatatgtcaaatgagttgttctcttggtatt 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 ACTAATTGTGCACGAGATAGTTGGTTGTCTATATGTCAAATGAGTTGTTCTCTTGGTATT 1140

Qy    1141 tgctctaccatctctccctagagcactctgtgtctatcccagtggataatttcccagttt 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 TGCTCTACCATCTCTCCCTAGAGCACTCTGTGTCTATCCCAGTGGATAATTTCCCAGTTT 1200

Qy    1201 actggtgatgattaggaaggttgttgatggttaggctaacctgccctggcccaaagccag 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 ACTGGTGATGATTAGGAAGGTTGTTGATGGTTAGGCTAACCTGCCCTGGCCCAAAGCCAG 1260

Qy    1261 acatgtacaagggctttctgtgagcaatgataagatctttgaatccaagatgcccagatg 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 ACATGTACAAGGGCTTTCTGTGAGCAATGATAAGATCTTTGAATCCAAGATGCCCAGATG 1320

Qy    1321 ttttaccagtcacaccctatggccatggctatacttggaagttctccttgttggcacaga 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 TTTTACCAGTCACACCCTATGGCCATGGCTATACTTGGAAGTTCTCCTTGTTGGCACAGA 1380

Qy    1381 catagaaatgctttaaccccaagcctttatatgggggacttctagctttgtgtcttgttt 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 CATAGAAATGCTTTAACCCCAAGCCTTTATATGGGGGACTTCTAGCTTTGTGTCTTGTTT 1440

Qy    1441 cagaccatgtggaatgataaatactcttttttgtgcttctgatctatcgatttcactaaca 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 CAGACCATGTGGAATGATAAATACTCTTTTTGTGCTTCTGATCTATCGATTTCACTAACA 1500

Qy    1501 tataccaagtaggtgctttgaaccccttttctgtaggctcacaccttaatctcaggcccct 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 TATACCAAGTAGGTGCTTTGAACCCCTTTCTGTAGGCTCACACCTTAATCTCAGGCCCCT 1560

Qy    1561 atatagtcacactttgatttaagaaaaacggagc 1594
          |||||||||||||||||||||||||||||||||
Db    1561 ATATAGTCACACTTTGATTTAAGAAAAACGGAGC 1594
```

```
RESULT   3
AC080033
LOCUS       AC080033              141262 bp    DNA     linear   HTG 13-FEB-2002
DEFINITION  Homo sapiens chromosome 8 clone RP11-885J16 map 8, *** SEQUENCING
            IN PROGRESS ***, 1 ordered pieces.
ACCESSION   AC080033
VERSION     AC080033.9  GI:18653568
KEYWORDS    HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      human.
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 141262)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-885J16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 141262)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
            FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
            Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,
            Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G.,
            Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
            O'Donnell,P., O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,
            Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-SEP-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 13, 2002 this sequence version replaced gi:18642748.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: Ll0939
                Center clone name: 885_J_16
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   141262: contig of 141262 bp in length.
FEATURES            Location/Qualifiers
     source         1. .141262
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="8"
                    /map="8"
                    /clone="RP11-885J16"
                    /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    44145 a  26647 c  26368 g  44102 t
ORIGIN
```

```
  Query Match             71.6%;  Score 1141.6;  DB 2;  Length 141262;
  Best Local Similarity   99.7%;  Pred. No. 2.7e-282;
  Matches 1144;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy    447 gtgatagcagggattagggaaactgaagagaaattctactacatcgtgcaggaagagaag 506
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  55935 GTGATAGCAGGGATTAGGGAAACTGAAGAGAAATTCTACTACATCGTGCAGGAAGAGAAG 55994

Qy    507 aactacagggaatccctaacccactgcaggattcggggtggaatgctagccatgcccaag 566
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  55995 AACTACAGGGAATCCCTAACCCACTGCAGGATTCGGGGTGGAATGCTAGCCATGCCCAAG 56054

Qy    567 gatgaagctgccaacacactcatcgctgactatgttgccaagagtggcttctttcgggtg 626
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  56055 GATGAAGCTGCCAACACACTCATCGCTGACTATGTTGCCAAGAGTGGCTTCTTTCGGGTG 56114

Qy    627 ttcattggcgtgaatgaccttgaaagggagggacagtacatgttcacagacaacactcca 686
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  56115 TTCATTGGCGTGAATGACCTTGAAAGGGAGGGACAGTACATGTTCACAGACAACACTCCA 56174

Qy    687 ctgcagaactatagcaactggaatgaggggggaacccagcgacccctatggtcatgaggac 746
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  56175 CTGCAGAACTATAGCAACTGGAATGAGGGGGAACCCAGCGACCCCTATGGTCATGAGGAC 56234

Qy    747 tgtgtggagatgctgagctctggcagatggaatgacacagagtgccatcttaccatgtac 806
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  56235 TGTGTGGAGATGCTGAGCTCTGGCAGATGGAATGACACAGAGTGCCATCTTACCATGTAC 56294

Qy    807 tttgtctgtgagttcatcaagaagaaaaagtaacttccctcatcctacgtatttgctatt 866
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  56295 TTTGTCTGTGAGTTCATCAAGAAGAAAAAGTAACTTCCCTCATCCTACGTATTTGCTATT 56354

Qy    867 ttcctgtgaccgtcattacagttattgttatccatccttttttttcctgattgtactacat 926
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  56355 TTCCTGTGACCGTCATTACAGTTATTGTTATCCATCCTTTTTTTTCCTGATTGTACTACAT 56414

Qy    927 ttgatctgagtcaacatagctagaaaatgctaaactgaggtatggagcctccatcatcat 986
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  56415 TTGATCTGAGTCAACATAGCTAGAAAATGCTAAACTGAGGTATGGAGCCTCCATCATCAT 56474

Qy    987 gctcttttgtgatgattttcatattttcacacatggtatgttattgacccaataactcgc 1046
          ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db  56475 GCTCTTTTGTGATGATTTTCATATTTTCACACATGGTATATTATTGACCCAATAACTCGC 56534

Qy   1047 caggttacatgggtcttgagagagaattttaattactaattgtgcacgagatagttggtt 1106
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  56535 CAGGTTACATGGGTCTTGAGAGAGAATTTTAATTACTAATTGTGCACGAGATAGTTGGTT 56594

Qy   1107 gtctatatgtcaaatgagttgttctcttggtatttgctctaccatctctccctagagcac 1166
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  56595 GTCTATATGTCAAATGAGTTGTTCTCTTGGTATTTGCTCTACCATCTCTCCCTAGAGCAC 56654

Qy   1167 tctgtgtctatcccagtggataatttcccagtttactggtgatgattaggaaggttgttg 1226
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  56655 TCTGTGTCTATCCCAGTGGATAATTTCCCAGTTTACTGGTGATGATTAGGAAGGTTGTTG 56714

Qy   1227 atggttaggctaacctgccctggcccaaagccagacatgtacaagggctttctgtgagca 1286
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  56715 ATGGTTAGGCTAACCTGCCCTGGCCCAAAGCCAGACATGTACAAGGGCTTTCTGTGAGCA 56774

Qy   1287 atgataagatctttgaatccaagatgcccagatgttttaccagtcacaccctatggccat 1346
          ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db  56775 ATGATAAGATCTTTGAATCCAAGATGCCCAGATCTTTTACCAGTCACACCCTATGGCCAT 56834

Qy   1347 ggctatacttggaagttctccttgttggcacagacatagaaatgctttaaccccaagcct 1406
          ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db  56835 GGCTATACTTGGAAGTTCTCCTTGTTGGCACAGACATAGAAATGCTTTAACCCCAAACCT 56894

Qy   1407 ttatatgggggacttctagctttgtgtcttgtttcagaccatgtggaatgataaatactc 1466
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  56895 TTATATGGGGGACTTCTAGCTTTGTGTCTTGTTTCAGACCATGTGGAATGATAAATACTC 56954

Qy   1467 tttttgtgcttctgatctatcgatttcactaacatataccaagtaggtgctttgaacccc 1526
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  56955 TTTTTGTGCTTCTGATCTATCGATTTCACTAACATATACCAAGTAGGTGCTTTGAACCCC 57014

Qy   1527 tttctgtaggctcacaccttaatctcaggcccctatatagtcacactttgatttaagaaa 1586
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  57015 TTTCTGTAGGCTCACACCTTAATCTCAGGCCCCTATATAGTCACACTTTGATTTAAGAAA 57074
```

```
Qy   1587 aacggagc 1594
          || |||||
Db  57075 AATGGAGC 57082


RESULT  4
AC097055/c
LOCUS       AC097055               221929 bp    DNA     linear   HTG 24-NOV-2001
DEFINITION  Rattus norvegicus clone CH230-2F22, WORKING DRAFT SEQUENCE, 11
            unordered pieces.
ACCESSION   AC097055
VERSION     AC097055.2  GI:17063112
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 221929)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 221929)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 24, 2001 this sequence version replaced gi:15982448.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: TUUS
                Center clone name: CH230-2F22
            -------------- Summary Statistics
                Sequencing vector: Plasmid; M77789
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 216965 bases at least Q40
                Consensus quality: 217905 bases at least Q30
                Consensus quality: 218702 bases at least Q20
                Estimated insert size: 218215; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 7.5x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    46448: contig of 46448 bp in length
            *   46449    46548: gap of unknown length
            *   46549    86887: contig of 40339 bp in length
            *   86888    86987: gap of unknown length
            *   86988   114267: contig of 27280 bp in length
            *  114268   114367: gap of unknown length
            *  114368   133628: contig of 19261 bp in length
            *  133629   133728: gap of unknown length
            *  133729   149817: contig of 16089 bp in length
            *  149818   149917: gap of unknown length
            *  149918   169202: contig of 19285 bp in length
            *  169203   169302: gap of unknown length
            *  169303   184404: contig of 15102 bp in length
            *  184405   184504: gap of unknown length
            *  184505   197588: contig of 13084 bp in length
            *  197589   197688: gap of unknown length
            *  197689   208937: contig of 11249 bp in length
            *  208938   209037: gap of unknown length
            *  209038   218041: contig of 9004 bp in length
            *  218042   218141: gap of unknown length
            *  218142   221929: contig of 3788 bp in length.
FEATURES             Location/Qualifiers
     source          1. .221929
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /clone="CH230-2F22"
BASE COUNT    65912 a  44454 c  45320 g  65130 t   1113 others
ORIGIN


  Query Match           21.5%;  Score 343;  DB 2;  Length 221929;
  Best Local Similarity  83.1%;  Pred. No. 2.9e-77;
  Matches  403;  Conservative    0;  Mismatches   80;  Indels    2;  Gaps    1;

Qy     447 gtgatagcagggattagggaaactgaagagaaattctactacatcgtgcaggaagagaag 506
           || |||||||||||  |||||||||||||||||||||||||||||| |||||||| || |||
Db   73194 GTCATAGCAGGGATCCGGGAAACTGAAGAGAAATTCTACTACATTGTGCAGGAGGAAAAG 73135

Qy     507 aactacagggaatccctaacccactgcaggattcggggtggaatgctagccatgcccaag 566
           ||||||||||||||| || |||||||||||||| ||||| || ||||| |||||||| |||
Db   73134 AACTACAGGGAATCTCTGACCCACTGCAGGATCCGGGGAGGGATGCTGGCCATGCCTAAG 73075

Qy     567 gatgaagctgccaacacactcatcgctgactatgttgccaagagtggcttctttcgggtg 626
           ||||||| ||  ||||| || || |||||||||||| ||||||||||||||||  | |||
Db   73074 GATGAAGTTGTTAACACTCTTATTGCTGACTATGTCGCCAAGAGTGGCTTCTTCAGAGTG 73015

Qy     627 ttcattggcgtgaatgaccttgaaagggagggacagtacatgttcacagacaacactcca 686
           |||||||||||||||||||||| | |||||| || ||  ||||||||||||||||||||
Db   73014 TTCATTGGCGTGAATGACCTTGAGAAGGAGGGGCAATATGTGTTCACAGACAACACTCCA 72955
```

```
Qy    687 ctgcagaactatagcaactggaatgaggggggaacccagcgacccctatggtcatgaggac 746
          |||||||||| |||||||||| |||||||| || || |||||||||| |||||||||
Db  72954 TTGCAGAACTACAGCAACTGGAAGGAGGGGGGAGCCTAGTGACCCCTATGGCCATGAGGAC 72895

Qy    747 tgtgtggagatgctgagctctggcagatggaatgacacagagtgccatcttaccatgtac 806
          ||||||||||||||||||||||||| ||||| |||||||||| ||||||||||||||
Db  72894 TGTGTGGAGATGCTGAGCTCTGGCAGGTGGAACGACACAGAGTGTCATCTTACCATGTAT 72835

Qy    807 tttgtctgtgagttcatcaagaagaaaaagtaacttccctcatcctacgtatttgctatt 866
          |||||||||||||| |||||||||||| ||| |||| |||| ||| | || | |
Db  72834 TTTGTCTGTGAGTTTGTCAAGAAGAAAAATTAATTTCCTTCATGTTACACAGGCACTTTC 72775

Qy    867 ttcctgtgaccgtcattacagttattgttatccatccttttttttcctgattgtactacat 926
          | || | | |  || | ||||| | |||| ||| ||  || || | |  | |  |
Db  72774 TCCCAGGGTC--TCCTAGCAGTTACTTTTATTCATTGTTGCCTTTCTAACTTCATTGTAC 72717

Qy    927 ttgat 931
          ||| |
Db  72716 TTGCT 72712
```

```
RESULT   5
BC000078
LOCUS       BC000078               1257 bp    mRNA    linear   PRI 12-JUL-2001
DEFINITION  Homo sapiens, clone MGC:3279 IMAGE:3507377, mRNA, complete cds.
ACCESSION   BC000078
VERSION     BC000078.1  GI:12652660
KEYWORDS    MGC.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1257)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-NOV-2000) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Rachel Dickhoff, Jessica Fahey; Stephanie Ford, Julia
            Greene, Mark Ketteman and Anuradha Madan

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 7 Row: g Column: 20
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis.
FEATURES             Location/Qualifiers
     source          1. .1257
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="MGC:3279 IMAGE:3507377"
                     /tissue_type="Brain, neuroblastoma"
                     /clone_lib="NIH_MGC_19"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     CDS             83. .898
                     /codon_start=1
                     /product="Unknown (protein for MGC:3279)"
                     /protein_id="AAH00078.1"
                     /db_xref="GI:12652661"
                     /translation="MRGNLALVGVLISLAFLSLLPSGHPQPAGDDACSVQILVPGLKG
                     DAGEKGDKGAPGRPGRVGPTGEKGDMGDKGQKGSVGRHGKIGPIGSKGEKGDSGDIGP
                     PGPNGEPGLPCECSQLRKAIGEMDNQVSQLTSELKFIKNAVAGVRETESKIYLLVKEE
                     KRYADAQLSCQGRGGTLSMPKDEAANGLMAAYLAQAGLARVFIGINDLEKEGAFVYSD
                     HSPMRTFNKWRSGEPNNAYDEEDCVEMVASGGWNDVACHTTMYFMCEFDKENM"
BASE COUNT      306 a    324 c    395 g    232 t
ORIGIN
```

```
  Query Match            15.4%;  Score 246;  DB 9;  Length 1257;
  Best Local Similarity  58.4%;  Pred. No. 2.5e-52;
  Matches  429;  Conservative    0;  Mismatches  305;  Indels    0;  Gaps    0;

Qy    106 ctaccgctgaagtctgtgccacacacacaatttcaccaggacccaaaggagatgatggtg 165
          || || ||| | |||  |    || |   |    || || | |||||| |||| || |
Db    165 CTGGCGATGACGCCTGCTCTGTGCAGATCCTCGTCCCTGGCCTCAAAGGGGATGCGGGAG 224

Qy    166 aaaaaggagatccaggagaagagggaaagcatggcaaagtgggacgcatggggccgaaag 225
          | || |||||   ||| |     |||  || ||| | ||| || | || |||    ||||
Db    225 AGAAGGGAGACAAAGGCGCCCCCGGACGGCCTGGAAGAGTCGGCCCCACGGGAGAAAAAG 284

Qy    226 gaattaaaggagaactgggtgatatgggagatcggggcaatattggcaagactgggccca 285
          ||   | || ||    ||  | |  ||   |  |||   |  ||| || | ||| ||||
Db    285 GAGACATGGGGGACAAAGGACAGAAAGGCAGTGTGGGTCGTCATGGAAAAATTGGTCCCA 344

Qy    286 ttgggaagaagggtgacaaaggggaaaaaggtttgcttggaatacctggagaaaaaggca 345
          ||||    || ||||| ||||| ||    |||   | |||   |||||    || ||
Db    345 TTGGCTCTAAAGGTGAGAAAGGAGATTCCGGTGACATAGGACCCCCTGGTCCTAATGGAG 404

Qy    346 aagcaggtactgtctgtgattgtggaagataccggaaatttgttggacaactggatatta 405
          || ||||       ||||| ||  |       || ||     | ||  |  |||| |
Db    405 AACCAGGCCTCCCATGTGAGTGCAGCCAGCTGCGCAAGGCCATCGGGGAGATGGACAACC 464

Qy    406 gtattgcccggctcaagacatctatgaagtttgtcaagaatgtgatagcagggattaggg 465
          |  | | ||| |  |       | |||||  |||||||||   | || ||  |  | |
Db    465 AGGTCTCTCAGCTGACCAGCGAGCTCAAGTTCATCAAGAATGCTGTCGCCGGTGTGCGCG 524

Qy    466 aaactgaagagaaattctactacatcgtgcaggaagagaagaactacagggaatccctaa 525
          | || ||    ||  |||||    | ||| |||| ||||||  ||||  |||  |||
Db    525 AGACGGAGAGCAAGATCTACCTGCTGGTGAAGGAGGAGAAGCGCTACGCGGACGCCCAGC 584

Qy    526 cccactgcaggattcggggtggaatgctagccatgcccaaggatgaagctgccaacacac 585
            ||||  |   || || || | |||   ||||||||||| || ||||||||    |
Db    585 TGTCCTGCCAGGGCCGCGGGGGCACGCTGAGCATGCCCAAGGACGAGGCTGCCAATGGCC 644

Qy    586 tcatcgctgactatgttgccaagagtggcttctttcgggtgttcattggcgtgaatgacc 645
          | || || |  ||  | ||  |    ||| |    || || ||||| ||| | || ||||
Db    645 TGATGGCCGCATACCTGGCGCAAGCCGGCCTGGCCCGTGTCTTCATCGGCATCAACGACC 704

Qy    646 ttgaaagggagggacagtacatgttcacagacaacactccactgcagaactatagcaact 705
          | || | ||||||    | | ||| | | ||| || | ||  ||| || ||  | ||| |
Db    705 TGGAGAAGGAGGGCGCCTTCGTGTACTCTGACCACTCCCCCATGCGGACCTTCAACAAGT 764

Qy    706 ggaatgaggggggaacccagcgacccctatggtcatgaggactgtgtggagatgctgagct 765
          ||      || || |||| | |  |||| |   | |||||||| |||||||| ||  ||
Db    765 GGCGCAGCGGTGAGCCCAACAATGCCTACGACGAGGAGGACTGCGTGGAGATGGTGGCCT 824

Qy    766 ctggcagatggaatgacacagagtgccatcttaccatgtactttgtctgtgagttcatca 825
          | ||| | ||||| |||   |  |||||    ||||||||||  | ||||||||   ||
Db    825 CGGGCGGCTGGAACGACGTGGCCTGCCACACCACCATGTACTTCATGTGTGAGTTTGACA 884

Qy    826 agaagaaaaagtaa 839
          || |||| | || |
Db    885 AGGAGAACATGTGA 898
```

```
RESULT   6
AC107953
LOCUS       AC107953               62064 bp    DNA     linear   HTG 24-JAN-2002
DEFINITION  Homo sapiens chromosome 8 clone RP11-278I4 map 8, LOW-PASS SEQUENCE
            SAMPLING.
ACCESSION   AC107953
VERSION     AC107953.1  GI:18308681
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 62064)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-278I4
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 62064)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L24493
                Center clone name: 278_I_4
            --------------.
            * NOTE: This record contains 75 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *      1     726: contig of 726 bp in length
            *    727 826: gap of      100 bp
            *    827    1539: contig of 713 bp in length
            *   1540 1639: gap of      100 bp
            *   1640    2385: contig of 746 bp in length
            *   2386 2485: gap of      100 bp
            *   2486    3214: contig of 729 bp in length
            *   3215 3314: gap of      100 bp
            *   3315    4069: contig of 755 bp in length
            *   4070 4169: gap of      100 bp
            *   4170    4914: contig of 745 bp in length
            *   4915 5014: gap of      100 bp
            *   5015    5738: contig of 724 bp in length
            *   5739 5838: gap of      100 bp
            *   5839    6555: contig of 717 bp in length
            *   6556 6655: gap of      100 bp
            *   6656    7383: contig of 728 bp in length
            *   7384 7483: gap of      100 bp
            *   7484    8213: contig of 730 bp in length
            *   8214 8313: gap of      100 bp
            *   8314    9034: contig of 721 bp in length
            *   9035 9134: gap of      100 bp
            *   9135    9866: contig of 732 bp in length
            *   9867 9966: gap of      100 bp
            *   9967   10690: contig of 724 bp in length
            *  10691 10790: gap of      100 bp
            *  10791   11531: contig of 741 bp in length
            *  11532 11631: gap of      100 bp
            *  11632   12373: contig of 742 bp in length
            *  12374 12473: gap of      100 bp
            *  12474   13215: contig of 742 bp in length
            *  13216 13315: gap of      100 bp
            *  13316   14031: contig of 716 bp in length
            *  14032 14131: gap of      100 bp
            *  14132   14833: contig of 702 bp in length
            *  14834 14933: gap of      100 bp
            *  14934   15665: contig of 732 bp in length
            *  15666 15765: gap of      100 bp
            *  15766   16512: contig of 747 bp in length
            *  16513 16612: gap of      100 bp
            *  16613   17353: contig of 741 bp in length
            *  17354 17453: gap of      100 bp
            *  17454   18194: contig of 741 bp in length
            *  18195 18294: gap of      100 bp
            *  18295   19033: contig of 739 bp in length
            *  19034 19133: gap of      100 bp
            *  19134   19852: contig of 719 bp in length
            *  19853 19952: gap of      100 bp
            *  19953   20676: contig of 724 bp in length
            *  20677 20776: gap of      100 bp
            *  20777   21522: contig of 746 bp in length
            *  21523 21622: gap of      100 bp
            *  21623   22357: contig of 735 bp in length
            *  22358 22457: gap of      100 bp
            *  22458   23175: contig of 718 bp in length
            *  23176 23275: gap of      100 bp
            *  23276   24007: contig of 732 bp in length
            *  24008 24107: gap of      100 bp
            *  24108   24837: contig of 730 bp in length
            *  24838 24937: gap of      100 bp
            *  24938   25672: contig of 735 bp in length
            *  25673 25772: gap of      100 bp
            *  25773   26512: contig of 740 bp in length
            *  26513 26612: gap of      100 bp
            *  26613   27356: contig of 744 bp in length
            *  27357 27456: gap of      100 bp
            *  27457   28206: contig of 750 bp in length
            *  28207 28306: gap of      100 bp
            *  28307   29013: contig of 707 bp in length
            *  29014 29113: gap of      100 bp
            *  29114   29836: contig of 723 bp in length
            *  29837 29936: gap of      100 bp
            *  29937   30669: contig of 733 bp in length
            *  30670 30769: gap of      100 bp
            *  30770   31497: contig of 728 bp in length
            *  31498 31597: gap of      100 bp
            *  31598   32322: contig of 725 bp in length
            *  32323 32422: gap of      100 bp
            *  32423   33148: contig of 726 bp in length
            *  33149 33248: gap of      100 bp
            *  33249   33988: contig of 740 bp in length
            *  33989 34088: gap of      100 bp
            *  34089   34818: contig of 730 bp in length
            *  34819 34918: gap of      100 bp
            *  34919   35644: contig of 726 bp in length
            *  35645 35744: gap of      100 bp
            *  35745   36475: contig of 731 bp in length
            *  36476 36575: gap of      100 bp
            *  36576   37300: contig of 725 bp in length
            *  37301 37400: gap of      100 bp
```

```
            *    37401    38119: contig of 719 bp in length
            *    38120 38219: gap of      100 bp
            *    38220    38939: contig of 720 bp in length
            *    38940 39039: gap of      100 bp
            *    39040    39765: contig of 726 bp in length
            *    39766 39865: gap of      100 bp
            *    39866    40598: contig of 733 bp in length
            *    40599 40698: gap of      100 bp
            *    40699    41434: contig of 736 bp in length
            *    41435 41534: gap of      100 bp
            *    41535    42273: contig of 739 bp in length
            *    42274 42373: gap of      100 bp
            *    42374    43068: contig of 695 bp in length
            *    43069 43168: gap of      100 bp
            *    43169    43917: contig of 749 bp in length
            *    43918 44017: gap of      100 bp
            *    44018    44723: contig of 706 bp in length
            *    44724 44823: gap of      100 bp
            *    44824    45555: contig of 732 bp in length
            *    45556 45655: gap of      100 bp
            *    45656    46351: contig of 696 bp in length
            *    46352 46451: gap of      100 bp
            *    46452    47173: contig of 722 bp in length
            *    47174 47273: gap of      100 bp
            *    47274    47992: contig of 719 bp in length
            *    47993 48092: gap of      100 bp
            *    48093    48811: contig of 719 bp in length
            *    48812 48911: gap of      100 bp
            *    48912    49629: contig of 718 bp in length
            *    49630 49729: gap of      100 bp
            *    49730    50468: contig of 739 bp in length
            *    50469 50568: gap of      100 bp
            *    50569    51287: contig of 719 bp in length
            *    51288 51387: gap of      100 bp
            *    51388    52114: contig of 727 bp in length
            *    52115 52214: gap of      100 bp
            *    52215    52955: contig of 741 bp in length
            *    52956 53055: gap of      100 bp
            *    53056    53797: contig of 742 bp in length
            *    53798 53897: gap of      100 bp
            *    53898    54629: contig of 732 bp in length
            *    54630 54729: gap of      100 bp
            *    54730    55450: contig of 721 bp in length
            *    55451 55550: gap of      100 bp
            *    55551    56279: contig of 729 bp in length
            *    56280 56379: gap of      100 bp
            *    56380    57106: contig of 727 bp in length

  Query Match              9.9%;  Score 158.6;  DB 2;  Length 62064;
  Best Local Similarity   92.3%;  Pred. No. 6.6e-30;
  Matches  167;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy      1 cagcaatgaatggctttgcatccttgcttcgaagaaaccaatttatcctcctggtactat 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  61465 CAGCAATGAATGGCTTTGCATCCTTGCTTCGAAGAAACCAATTTATCCTCCTGGTACTAT 61524

Qy     61 ttcttttgcaaattcagagtctgggtctggatattgatagccgtcctaccgctgaagtct 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  61525 TTCTTTTGCAAATTCAGAGTCTGGGTCTGGATATTGATAGCCGTCCTACCGCTGAAGTCT 61584

Qy    121 gtgccacacacacaatttcaccaggacccaaaggagatgatggtgaaaaaggagatccag 180
          |||||||||||||||||||||||||||||||| || ||  |  ||  |  |  | ||
Db  61585 GTGCCACACACACAATTTCACCAGGACCCAAAGGTGAGGAAAGAAAACCACAATTTTCAT 61644

Qy    181 g 181
          |
Db  61645 G 61645


RESULT   7
AC023487
LOCUS       AC023487              182475 bp    DNA     linear   HTG 26-MAR-2001
DEFINITION  Homo sapiens chromosome 8 clone RP11-164H21, WORKING DRAFT
            SEQUENCE, 3 unordered pieces.
ACCESSION   AC023487
VERSION     AC023487.10  GI:13357236
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 182475)
  AUTHORS   Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
            Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
            Mao,J., Komp,C., Kottler,S., Lam,B., Marathe,R., Miranda,M.,
            Morehouse,A.J., Nguyen,M., Oefner,P., Palm,C.J., Ramirez,D.,
            Southwick,A.M., Webb,C., Wilhelmy,J., Yu,S. and Davis,R.W.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 182475)
  AUTHORS   Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D., Federspiel,N.,
            Glukhov,S., Hansen,N., Hyman,R., Mao,J., Marathe,R.,
            Morehouse,A.J., Oefner,P., Palm,C.J., Ramirez,D., Wilhelmy,J.,
            Yu,S. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-FEB-2000) DNA Sequencing and Technology Center,
            Stanford University, 855 California Avenue, Palo Alto, CA 94304,
            USA
COMMENT     On Mar 16, 2001 this sequence version replaced gi:13324778.
            -------------- Genome Center
                 Center: Stanford DNA Sequencing and Technology Development
            Center
                 Center code: SDSTDC
                 Web site: http://sequence-www.stanford.edu/group/human/
                 Contact: hum-info@sequence.stanford.edu
            -------------- Project Information
                 Center project name: 844
                 Center clone name: RP11-164H21
            -------------- Summary Statistics
                 Sequencing Vector: M13mp18; X02513; 98% of reads
                 Sequencing Vector: plasmid; plasmid_accession; 2% of reads
                 Chemistry: Dye-primer; 0% of reads
                 Chemistry: Dye-terminator Big Dye; 99% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 180260 bases at least Q40
                 Consensus quality: 180441 bases at least Q30
                 Consensus quality: 180507 bases at least Q20
                 Insert size: 178614; agarose-fp
                 Insert size: 182275; sum-of-contigs
                 Quality coverage:  8.1x in Q20 bases; agarose-fp
                 Quality coverage:  7.9x in Q20 bases; sum-of-contigs.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    42072: contig of 42072 bp in length
            *    42073    42172: gap of unknown length
            *    42173   109254: contig of 67082 bp in length
            *   109255   109354: gap of unknown length
            *   109355   182475: contig of 73121 bp in length.
FEATURES             Location/Qualifiers
     source          1. .182475
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /clone="RP11-164H21"
                     /clone_lib="RPCI human BAC library 11"
     misc_feature    1. .42072
                     /note="assembly_name:Contig9
                     clone_end:T7"
     misc_feature    42173. .109254
                     /note="assembly_name:Contig10"
     misc_feature    109355. .182475
```

```
                     /note="assembly_name:Contig11
                     clone_end:SP6"
BASE COUNT     56447 a  33954 c  34262 g  57610 t    202 others
ORIGIN


  Query Match             9.9%;  Score 158.6;  DB 2;  Length 182475;
  Best Local Similarity   92.3%;  Pred. No. 6.5e-30;
  Matches  167;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy       1 cagcaatgaatggctttgcatccttgcttcgaagaaaccaatttatcctcctggtactat 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 105630 CAGCAATGAATGGCTTTGCATCCTTGCTTCGAAGAAACCAATTTATCCTCCTGGTACTAT 105689

Qy      61 ttcttttgcaaattcagagtctgggtctggatattgatagccgtcctaccgctgaagtct 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 105690 TTCTTTTGCAAATTCAGAGTCTGGGTCTGGATATTGATAGCCGTCCTACCGCTGAAGTCT 105749

Qy     121 gtgccacacacacaatttcaccaggacccaaaggagatgatggtgaaaaaggagatccag 180
           |||||||||||||||||||||||||||||||| || ||  |  ||  |  |  | ||
Db 105750 GTGCCACACACACAATTTCACCAGGACCCAAAGGTGAGGAAAGAAAACCACAATTTTCAT 105809

Qy     181 g 181
           |
Db 105810 G 105810


RESULT   8
AC010907
LOCUS       AC010907              139357 bp    DNA     linear   PRI 09-JAN-2002
DEFINITION  Homo sapiens BAC clone RP11-568H24 from 2, complete sequence.
ACCESSION   AC010907
VERSION     AC010907.10  GI:15321567
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 139357)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
REFERENCE   2  (bases 1 to 139357)
  AUTHORS   Du,H., Haakenson,W. and Dixon,R.
  TITLE     The sequence of Homo sapiens BAC clone RP11-568H24
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 139357)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-SEP-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 139357)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-AUG-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 139357)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Aug 28, 2001 this sequence version replaced gi:13399437.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
               Center project name: H_NH0568H24
            --------------

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO.  For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries.  Genomics 51:1-8.  The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at the Roswell Park Cancer Institute
            (http://bacpac.med.buffalo.edu)
            VECTOR:  pBACe3.6
            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the right is RP11-178E6, 2000 bp overlap.
            Actual start of this clone is at base position 1 of RP11-568H24.

            The sequence between 66093 to 66578 and 104506 to 104590 is covered
            only by pcr products from clone DNA.  he sequence contains a
            dinucleotide (TG) run from 65513 to 65634 in which the exact length
            is unknown.  The sequence contains a dinucleotide (TC) run from
            104386 bp to 104631 bp in which the exact length is unknown.  The
            sequence from base position 4458 to 6187 can not be guaranted due
            to  a tandem repeat.
FEATURES             Location/Qualifiers
     source          1. .139357
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2"
                     /clone="RP11-568H24"
                     /clone_lib="RPCI-11"
     repeat_region   383. .409
                     /rpt_family="AT_rich"
     repeat_region   876. .950
                     /rpt_family="AcHobo"
     repeat_region   951. .1248
                     /rpt_family="Alu"
     repeat_region   1249. .1339
                     /rpt_family="AcHobo"
     repeat_region   1340. .1514
                     /rpt_family="MER1_type"
     repeat_region   1611. .1965
                     /rpt_family="AcHobo"
     repeat_region   1973. .2279
                     /rpt_family="Alu"
     repeat_region   2324. .2843
                     /rpt_family="L1"
     repeat_region   2844. .3050
                     /rpt_family="MER1_type"
     misc_feature    2943. .2958
                     /note="similar to Homo sapiens EST AI597790 (NID:g4606838)
                     tr92h02.x1"
     repeat_region   3051. .3174
                     /rpt_family="L1"
     repeat_region   3175. .3483
```

```
                     /rpt_family="Alu"
misc_feature         3182. .3206
                     /note="similar to Homo sapiens EST BG178637
                     (NID:g12685340)"
repeat_region        3484. .4227
                     /rpt_family="L1"
repeat_region        4255. .4330
                     /rpt_family="(T)n"
repeat_region        4312. .4987
                     /rpt_family="SVA"
misc_feature         4571. .6524
                     /note="CpG_island (%GC=74.6, o/e=0.80, #CpGs=159)"
repeat_region        4978. .5152
                     /rpt_family="(CCGG)n"
repeat_region        5163. .5560
                     /rpt_family="SVA"
repeat_region        5249. .5273
                     /rpt_family="GC_rich"
repeat_region        5593. .6486
                     /rpt_family="SVA"
repeat_region        6443. .6529
                     /rpt_family="GA-rich"
repeat_region        6531. .6732
                     /rpt_family="L1"
repeat_region        6733. .7033
                     /rpt_family="Alu"
repeat_region        7034. .7249
                     /rpt_family="L1"
repeat_region        7250. .7554
                     /rpt_family="Alu"
repeat_region        7555. .8490
                     /rpt_family="L1"
repeat_region        8528. .8726
                     /rpt_family="AcHobo"
misc_feature         8666. .8669
                     /note="match to EST BF314952 (NID:g11263123)"
misc_feature         9031. .9360
                     /note="match to EST BF314952 (NID:g11263123)"
misc_feature         9047. .9360
                     /note="match to EST BE259985 (NID:g9130920)"
misc_feature         9068. .9069
                     /note="match to EST BF316717 (NID:g11265074)"
misc_feature         9068. .9069
                     /note="match to EST BI198782 (NID:g14653803)"
misc_feature         9144. .9147
                     /note="similar to Homo sapiens EST BF311850
                     (NID:g11259616)"
misc_feature         9239. .9360
                     /note="similar to Homo sapiens EST BI198308
                     (NID:g14653329)"
misc_feature         9239. .9360
                     /note="match to EST BI198831 (NID:g14653852)"
misc_feature         9239. .9360
                     /note="similar to Homo sapiens EST BI198502
                     (NID:g14653523)"
misc_feature         9240. .9360
                     /note="match to EST BF311988 (NID:g11259763)"
misc_feature         9240. .9360
                     /note="match to EST BF206235 (NID:g11099821)"
misc_feature         9240. .9360
                     /note="similar to Homo sapiens EST BF313026
                     (NID:g11260952)"
misc_feature         9240. .9360
                     /note="match to EST BI199896 (NID:g14654917)"
misc_feature         9313. .9360
                     /note="match to EST BF311981 (NID:g11259754)"
misc_feature         9313. .9360
                     /note="similar to Homo sapiens EST BF316085
                     (NID:g11264380)"
misc_feature         9313. .9360
                     /note="similar to Homo sapiens EST BI198293
                     (NID:g14653314)"
misc_feature         9322. .9360
                     /note="similar to Homo sapiens EST BF311850
                     (NID:g11259616)"
misc_feature         9322. .9360
                     /note="match to EST BF316717 (NID:g11265074)"
misc_feature         9322. .9360
                     /note="match to EST BI198782 (NID:g14653803)"
misc_feature         9333. .9360
                     /note="match to EST BE313758 (NID:g9134050)"
misc_feature         9338. .9360
                     /note="match to EST BE260359 (NID:g9131557)"

  Query Match             9.8%;  Score 157;  DB 9;  Length 139357;
  Best Local Similarity   62.7%;  Pred. No. 1.7e-29;
  Matches  244;  Conservative    0;  Mismatches  145;  Indels    0;  Gaps    0;

Qy     451 tagcagggattagggaaactgaagagaaattctactacatcgtgcaggaagagaagaact 510
           | || ||  |  | || || ||    ||  |||||    | ||| |||| ||||||  ||
Db   57922 TCGCCGGTGTGCGCGAGACGGAGAGCAAGATCTACCTGCTGGTGAAGGAGGAGAAGCGCT 57981

Qy     511 acagggaatccctaacccactgcaggattcggggtggaatgctagccatgcccaaggatg 570
           ||  |||  |||       ||||  |   || || || | |||   ||||||||||||| |
Db   57982 ACGCGGACGCCCAGCTGTCCTGCCAGGGCCGCGGGGGCACGCTGAGCATGCCCAAGGACG 58041

Qy     571 aagctgccaacacactcatcgctgactatgttgccaagagtggcttctttcgggtgttca 630
           | ||||||||    || || || |  ||  | ||  |    ||| |    || || ||||
Db   58042 AGGCTGCCAATGGCCTGATGGCCGCATACCTGGCGCAAGCCGGCCTGGCCCGTGTCTTCA 58101

Qy     631 ttggcgtgaatgaccttgaaagggagggacagtacatgttcacagacaacactccactgc 690
           | ||| | || |||||| || | ||||||    | | ||| | | ||| || | ||  |||
Db   58102 TCGGCATCAACGACCTGGAGAAGGAGGGCGCCTTCGTGTACTCTGACCACTCCCCCATGC 58161

Qy     691 agaactatagcaactggaatgaggggaacccagcgacccctatggtcatgaggactgtg 750
           || ||  | ||| |||      || || |||| | |  |||| |    | |||||||| |
Db   58162 GGACCTTCAACAAGTGGCGCAGCGGTGAGCCCAACAATGCCTACGACGAGGAGGACTGCG 58221

Qy     751 tggagatgctgagctctggcagatggaatgacacagagtgccatcttaccatgtactttg 810
           |||||||| ||  ||| ||| | ||||| |||    |  |||||    ||||||||||
Db   58222 TGGAGATGGTGGCCTCGGGCGGCTGGAACGACGTGGCCTGCCACACCACCATGTACTTCA 58281

Qy     811 tctgtgagttcatcaagaagaaaaagtaa 839
           | ||||||||   |||| |||| | || |
Db   58282 TGTGTGAGTTTGACAAGGAGAACATGTGA 58310


RESULT   9
AC108488/c
LOCUS       AC108488              204511 bp    DNA     linear   HTG 09-FEB-2002
DEFINITION  Homo sapiens chromosome 2 clone RP13-512J5, WORKING DRAFT SEQUENCE,
            10 unordered pieces.
ACCESSION   AC108488
VERSION     AC108488.2  GI:18642889
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 204511)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 204511)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JAN-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Feb 9, 2002 this sequence version replaced gi:18390293.

            ------------------- Genome Center --------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
```

```
            ---------- Project Information ----------
            Center project name: H_FH0512J05
            ----------------------------------------------------
            ------------ Summary Statistics ------------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 99%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 99% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 189568 bases at least Q40
            Consensus quality: 191160 bases at least Q30
            Consensus quality: 202880 bases at least Q20
            Insert size: 199000; agarose-fp
            Insert size: 204630; sum-of-contigs
            Quality coverage: 6.54 in Q20 bases; agarose-fp
            Quality coverage: 6.57 in Q20 bases; sum-of-contigs
            ----------------------------------------------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1098: contig of 1098 bp in length
            *     1099     1198: gap of unknown length
            *     1199     2264: contig of 1066 bp in length
            *     2265     2364: gap of unknown length
            *     2365     3468: contig of 1104 bp in length
            *     3469     3568: gap of unknown length
            *     3569    13569: contig of 10001 bp in length
            *    13570    13669: gap of unknown length
            *    13670    16966: contig of 3297 bp in length
            *    16967    17066: gap of unknown length
            *    17067    25252: contig of 8186 bp in length
            *    25253    25352: gap of unknown length
            *    25353    37713: contig of 12361 bp in length
            *    37714    37813: gap of unknown length
            *    37814    47622: contig of 9809 bp in length
            *    47623    47722: gap of unknown length
            *    47723    67227: contig of 19505 bp in length
            *    67228    67327: gap of unknown length
            *    67328   204511: contig of 137184 bp in length.
FEATURES             Location/Qualifiers
     source          1. .204511
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /clone="RP13-512J5"
     misc_feature    1. .1098
                     /note="assembly_name:Contig8"
     misc_feature    1199. .2264
                     /note="assembly_name:Contig13"
     misc_feature    2365. .3468
                     /note="assembly_name:Contig24"
     misc_feature    3569. .13569
                     /note="assembly_name:Contig26"
     misc_feature    13670. .16966
                     /note="assembly_name:Contig30"
     misc_feature    17067. .25252
                     /note="assembly_name:Contig31"
     misc_feature    25353. .37713
                     /note="assembly_name:Contig32"
     misc_feature    37814. .47622
                     /note="assembly_name:Contig33"
     misc_feature    47723. .67227
                     /note="assembly_name:Contig34"
     misc_feature    67328. .204511
                     /note="assembly_name:Contig35"
BASE COUNT    56511 a  50158 c  47466 g  49475 t    901 others
ORIGIN
```

```
  Query Match             9.8%;  Score 157;  DB 2;  Length 204511;
  Best Local Similarity   62.7%;  Pred. No. 1.7e-29;
  Matches  244;  Conservative    0;  Mismatches  145;  Indels    0;  Gaps    0;

Qy     451 tagcagggattagggaaactgaagagaaatttctactacatcgtgcaggaagagaagaact 510
           | || ||  | | || || ||    ||  |||||    | ||| |||| ||||||  ||
Db 147804 TCGCCGGTGTGCGCGAGACGGAGAGCAAGATCTACCTGCTGGTGAAGGAGGAGAAGCGCT 147745

Qy     511 acagggaatccctaacccactgcaggattcggggtggaatgctagccatgcccaaggatg 570
           ||  |||  |||       ||||  |   || || || | |||   ||||||||||| |
Db 147744 ACGCGGACGCCCAGCTGTCCTGCCAGGGCCGCGGGGGCACGCTGAGCATGCCCAAGGACG 147685

Qy     571 aagctgccaacacactcatcgctgactatgttgccaagagtggcttctttcgggtgttca 630
           | ||||||||    || || || |  ||  | ||  |    ||| |    || || ||||
Db 147684 AGGCTGCCAATGGCCTGATGGCCGCATACCTGGCGCAAGCCGGCCTGGCCCGTGTCTTCA 147625

Qy     631 ttggcgtgaatgaccttgaaagggagggacagtacatgttcacagacaacactccactgc 690
           | ||| | || |||| || | ||||||    | | ||| | | ||| || | ||  |||
Db 147624 TCGGCATCAACGACCTGGAGAAGGAGGGCGCCTTCGTGTACTCTGACCACTCCCCCATGC 147565

Qy     691 agaactatagcaactggaatgaggggaacccagcgacccctatggtcatgaggactgtg 750
           || ||  | ||| |||      || || |||| | |  |||| |    | ||||||| |
Db 147564 GGACCTTCAACAAGTGGCGCAGCGGTGAGCCCAACAATGCCTACGACGAGGAGGACTGCG 147505

Qy     751 tggagatgctgagctctggcagatggaatgacacagagtgccatcttaccatgtactttg 810
           |||||||| ||  ||| ||| | ||||| |||   |  |||||    ||||||||||
Db 147504 TGGAGATGGTGGCCTCGGGCGGCTGGAACGACGTGGCCTGCCACACCACCATGTACTTCA 147445

Qy     811 tctgtgagttcatcaagaagaaaaagtaa 839
           | ||||||||   |||| |||| | || |
Db 147444 TGTGTGAGTTTGACAAGGAGAACATGTGA 147416


RESULT  10
AC106609
LOCUS       AC106609              105156 bp    DNA     linear   HTG 12-JAN-2002
DEFINITION  Rattus norvegicus clone CH230-171E21, *** SEQUENCING IN PROGRESS
            ***, 47 unordered pieces.
ACCESSION   AC106609
VERSION     AC106609.1  GI:18139133
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 105156)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
```

```
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 105156)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JAN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GLRH
                Center clone name: CH230-171E21
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329First call to
            findPhrapList
                Consensus quality: 86004 bases at least Q40
                Consensus quality: 92347 bases at least Q30
                Consensus quality: 98411 bases at least Q20
                Estimated insert size: 88922; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 1x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 47 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     9624: contig of 9624 bp in length
            *     9625     9724: gap of unknown length
            *     9725    15156: contig of 5432 bp in length
            *    15157    15256: gap of unknown length
            *    15257    19952: contig of 4696 bp in length
            *    19953    20052: gap of unknown length
            *    20053    25058: contig of 5006 bp in length
            *    25059    25158: gap of unknown length
            *    25159    27649: contig of 2491 bp in length
            *    27650    27749: gap of unknown length
            *    27750    30896: contig of 3147 bp in length
            *    30897    30996: gap of unknown length
            *    30997    34042: contig of 3046 bp in length
            *    34043    34142: gap of unknown length
            *    34143    36632: contig of 2490 bp in length
            *    36633    36732: gap of unknown length
            *    36733    38242: contig of 1510 bp in length
            *    38243    38342: gap of unknown length
            *    38343    40508: contig of 2166 bp in length
            *    40509    40608: gap of unknown length
            *    40609    42570: contig of 1962 bp in length
            *    42571    42670: gap of unknown length
            *    42671    45486: contig of 2816 bp in length
            *    45487    45586: gap of unknown length
            *    45587    47853: contig of 2267 bp in length
            *    47854    47953: gap of unknown length
            *    47954    49216: contig of 1263 bp in length
            *    49217    49316: gap of unknown length
            *    49317    51588: contig of 2272 bp in length
            *    51589    51688: gap of unknown length
            *    51689    52938: contig of 1250 bp in length
            *    52939    53038: gap of unknown length
            *    53039    56037: contig of 2999 bp in length
            *    56038    56137: gap of unknown length
            *    56138    57860: contig of 1723 bp in length
            *    57861    57960: gap of unknown length
            *    57961    60744: contig of 2784 bp in length
            *    60745    60844: gap of unknown length
            *    60845    61970: contig of 1126 bp in length
            *    61971    62070: gap of unknown length
            *    62071    63505: contig of 1435 bp in length
            *    63506    63605: gap of unknown length
            *    63606    65422: contig of 1817 bp in length
            *    65423    65522: gap of unknown length
            *    65523    68140: contig of 2618 bp in length
            *    68141    68240: gap of unknown length
            *    68241    70228: contig of 1988 bp in length
            *    70229    70328: gap of unknown length
            *    70329    71987: contig of 1659 bp in length
            *    71988    72087: gap of unknown length
            *    72088    73780: contig of 1693 bp in length
            *    73781    73880: gap of unknown length
            *    73881    75434: contig of 1554 bp in length
            *    75435    75534: gap of unknown length
            *    75535    77294: contig of 1760 bp in length
            *    77295    77394: gap of unknown length
            *    77395    78915: contig of 1521 bp in length
            *    78916    79015: gap of unknown length
            *    79016    80218: contig of 1203 bp in length
            *    80219    80318: gap of unknown length
            *    80319    82103: contig of 1785 bp in length
            *    82104    82203: gap of unknown length
            *    82204    83631: contig of 1428 bp in length
            *    83632    83731: gap of unknown length
            *    83732    85052: contig of 1321 bp in length
            *    85053    85152: gap of unknown length
            *    85153    86789: contig of 1637 bp in length
            *    86790    86889: gap of unknown length
            *    86890    88124: contig of 1235 bp in length
            *    88125    88224: gap of unknown length
            *    88225    89272: contig of 1048 bp in length
            *    89273    89372: gap of unknown length
            *    89373    90949: contig of 1577 bp in length
            *    90950    91049: gap of unknown length
            *    91050    92381: contig of 1332 bp in length
            *    92382    92481: gap of unknown length
            *    92482    94058: contig of 1577 bp in length
            *    94059    94158: gap of unknown length
            *    94159    95170: contig of 1012 bp in length
            *    95171    95270: gap of unknown length
            *    95271    96635: contig of 1365 bp in length
            *    96636    96735: gap of unknown length
            *    96736    98052: contig of 1317 bp in length
            *    98053    98152: gap of unknown length
            *    98153    99682: contig of 1530 bp in length
            *    99683    99782: gap of unknown length
            *    99783   101164: contig of 1382 bp in length
            *   101165   101264: gap of unknown length
            *   101265   102526: contig of 1262 bp in length
            *   102527   102626: gap of unknown length
            *   102627   103859: contig of 1233 bp in length
            *   103860   103959: gap of unknown length
            *   103960   105156: contig of 1197 bp in length.
FEATURES                Location/Qualifiers
     source             1. .105156
                        /organism="Rattus norvegicus"
```

```
                     /db_xref="taxon:10116"
                     /clone="CH230-171E21"
BASE COUNT    30932 a  19731 c  19545 g  30274 t   4674 others
ORIGIN


  Query Match             6.9%;  Score 110.8;  DB 2;  Length 105156;
  Best Local Similarity   80.2%;  Pred. No. 1.2e-17;
  Matches  130;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy      1 cagcaatgaatggctttgcatccttgcttcgaagaaaccaatttatcctcctggtactat 60
          |||  ||||||||||| |  |  | ||||||||||||||||| || ||||| ||| | || |
Db  32633 CAGTCATGAATGGCTGTAGAGTCCTGCTTCGAAGAAACCAGTTCATCCTGCTGCTGCTGT 32692

Qy     61 ttctttttgcaaattcagagtctgggtctggatattgatagccgtcctaccgctgaagtct 120
          |||| |||||  |||||||||||||||||||| ||||| | ||  || | || |||||||
Db  32693 TTCTCTTGCACTTTCAGAGTCTGGGTCTGGATGTTGATGGTCGCTCTGCTGCAGAAGTCT 32752

Qy    121 gtgccacacacacaatttcaccaggacccaaaggagatgatg 162
          ||| |||||| || |||||||||||||| ||||| || || |
Db  32753 GTGTCACACATACCATTTCACCAGGACCTAAAGGTGAGGAAG 32794


RESULT  11
HSU92547/c
LOCUS       HSU92547                 767 bp    DNA     linear   STS 26-OCT-1997
DEFINITION  Homo sapiens chromosome 8 STS, sequence tagged site.
ACCESSION   U92547
VERSION     U92547.1  GI:2564795
KEYWORDS    STS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 767)
  AUTHORS   Pierce,J., Leach,R. and Naylor,S.
  TITLE     New STS markers for human chromosome 8
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 767)
  AUTHORS   Pierce,J., Leach,R. and Naylor,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAR-1997) Pathology, UTHSCSA, 7703 Floyd Curl Drive,
            San Antonio, TX 78284, USA
FEATURES             Location/Qualifiers
     source          1. .767
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8 pter-qter"
                     /cell_type="hamster (CHO)/human lymphocyte UV20HL21-17
                     hybrid"
                     /clone_lib="LL08NS02 from Lawrence Livermore Laboratory"
                     /note="chromosome 8 flow sorted DNA"
BASE COUNT      223 a    146 c    115 g    252 t     31 others
ORIGIN


  Query Match             6.0%;  Score 95.4;  DB 11;  Length 767;
  Best Local Similarity   98.0%;  Pred. No. 1.2e-13;
  Matches   96;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy    350 aggtactgtctgtgattgtggaagataccggaaatttgttggacaactggatattagtat 409
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    540 AGGTACTGTCTGTGATTGTGGAAGATACCGGAAATTTGTTGGACAACTGGATATTAGTAT 481

Qy    410 tgcccggctcaagacatctatgaagtttgtcaagaatg 447
          ||| |||||||||||||||||||| ||||||||||
Db    480 TGCTCGGCTCAAGACATCTATGAAGTNTGTCAAGAATG 443


RESULT  12
I66494/c
LOCUS       I66494                   7218 bp    DNA     linear   PAT 28-DEC-1997
DEFINITION  Sequence 14 from patent US 5670367.
ACCESSION   I66494
VERSION     I66494.1  GI:2724471
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 7218)
  AUTHORS   Dorner,F., Scheiflinger,F. and Falkner,F.Gunter.
  TITLE     Recombinant fowlpox virus
  JOURNAL   Patent: US 5670367-A 14 23-SEP-1997;
FEATURES             Location/Qualifiers
     source          1. .7218
                     /organism="unknown"
BASE COUNT     1944 a   1491 c   1486 g   1929 t    368 others
ORIGIN


  Query Match             5.6%;  Score 89.8;  DB 6;  Length 7218;
  Best Local Similarity   2.6%;  Pred. No. 3.1e-12;
  Matches   10;  Conservative  253;  Mismatches  120;  Indels    0;  Gaps    0;

Qy    149 caaaggagatgatggtgaaaaaggagatccaggagaagagggaaagcatggcaaagtggg 208
          |::::::: :: :: :::::::::::   ::::::::::::::::: : :: :::: :::
Db   1435 CRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1376

Qy    209 acgcatggggccgaaaggaattaaaggagaactgggtgatatgggagatcggggcaatat 268
          : : : :::: :::::::: ::::::::: ::: :: : :::::: :::: :: :
Db   1375 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1316

Qy    269 tggcaagactgggcccattgggaagaagggtgacaaaggggaaaaaggtttgcttggaat 328
          :: :::: :::   :  :::::::::::: :: :::::::::::::   :   ::::
Db   1315 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1256

Qy    329 acctggagaaaaaggcaaagcaggtactgtctgtgattgtggaagataccggaaatttgt 388
          :   :::::::::: :::: ::: :  :   : ::  : :::::: :  :::::   :
Db   1255 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1196

Qy    389 tggacaactggatattagtattgcccggctcaagacatctatgaagtttgtcaagaatgt 448
          ::: ::  ::: :  :: :  :   ::   :::: :   : ::::   :  ::::: :
Db   1195 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1136

Qy    449 gatagcagggattagggaaactgaagagaaattctactacatcgtgcaggaagagaagaa 508
          :: :: :::::  :::::::  :::::::::    :  : :  : : :::::::::::::
Db   1135 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1076

Qy    509 ctacagggaatccctaacccact 531
            : :::::||| | | | | ||
Db   1075 RRRRRRRRRATCGCAAGCTCCCT 1053


RESULT  13
AC023487/c
LOCUS       AC023487                 182475 bp    DNA     linear   HTG 26-MAR-2001
DEFINITION  Homo sapiens chromosome 8 clone RP11-164H21, WORKING DRAFT
            SEQUENCE, 3 unordered pieces.
ACCESSION   AC023487
VERSION     AC023487.10  GI:13357236
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 182475)
  AUTHORS   Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
            Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
            Mao,J., Komp,C., Kottler,S., Lam,B., Marathe,R., Miranda,M.,
            Morehouse,A.J., Nguyen,M., Oefner,P., Palm,C.J., Ramirez,D.,
            Southwick,A.M., Webb,C., Wilhelmy,J., Yu,S. and Davis,R.W.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 182475)
  AUTHORS   Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D., Federspiel,N.,
            Glukhov,S., Hansen,N., Hyman,R., Mao,J., Marathe,R.,
```

```
            Morehouse,A.J., Oefner,P., Palm,C.J., Ramirez,D., Wilhelmy,J.,
            Yu,S. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-FEB-2000) DNA Sequencing and Technology Center,
            Stanford University, 855 California Avenue, Palo Alto, CA 94304,
            USA
COMMENT     On Mar 16, 2001 this sequence version replaced gi:13324778.
            -------------- Genome Center
                 Center: Stanford DNA Sequencing and Technology Development
            Center
                 Center code: SDSTDC
                 Web site: http://sequence-www.stanford.edu/group/human/
                 Contact: hum-info@sequence.stanford.edu
            -------------- Project Information
                 Center project name: 844
                 Center clone name: RP11-164H21
            -------------- Summary Statistics
                 Sequencing Vector: M13mp18; X02513; 98% of reads
                 Sequencing Vector: plasmid; plasmid_accession; 2% of reads
                 Chemistry: Dye-primer; 0% of reads
                 Chemistry: Dye-terminator Big Dye; 99% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 180260 bases at least Q40
                 Consensus quality: 180441 bases at least Q30
                 Consensus quality: 180507 bases at least Q20
                 Insert size: 178614; agarose-fp
                 Insert size: 182275; sum-of-contigs
                 Quality coverage:  8.1x in Q20 bases; agarose-fp
                 Quality coverage:  7.9x in Q20 bases; sum-of-contigs.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *      1    42072: contig of 42072 bp in length
            *  42073    42172: gap of unknown length
            *  42173   109254: contig of 67082 bp in length
            * 109255   109354: gap of unknown length
            * 109355   182475: contig of 73121 bp in length.
FEATURES             Location/Qualifiers
     source          1. .182475
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /clone="RP11-164H21"
                     /clone_lib="RPCI human BAC library 11"
     misc_feature    1. .42072
                     /note="assembly_name:Contig9
                     clone_end:T7"
     misc_feature    42173. .109254
                     /note="assembly_name:Contig10"
     misc_feature    109355. .182475
                     /note="assembly_name:Contig11
                     clone_end:SP6"
BASE COUNT    56447 a  33954 c  34262 g  57610 t    202 others
ORIGIN


  Query Match             5.0%;  Score 79;  DB 2;  Length 182475;
  Best Local Similarity   99.7%;  Pred. No. 1.8e-09;
  Matches 1144;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy    447 gtgatagcagggattagggaaactgaagagaaattctactacatcgtgcaggaagagaag 506
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6233 GTGATAGCAGGGATTAGGGAAACTGAAGAGAAATTCTACTACATCGTGCAGGAAGAGAAG 6174

Qy    507 aactacagggaatccctaacccactgcaggattcggggtggaatgctagccatgcccaag 566
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6173 AACTACAGGGAATCCCTAACCCACTGCAGGATTCGGGGTGGAATGCTAGCCATGCCCAAG 6114

Qy    567 gatgaagctgccaacacactcatcgctgactatgttgccaagagtggcttctttcgggtg 626
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6113 GATGAAGCTGCCAACACACTCATCGCTGACTATGTTGCCAAGAGTGGCTTCTTTCGGGTG 6054

Qy    627 ttcattggcgtgaatgaccttgaaagggagggacagtacatgttcacagacaacactcca 686
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6053 TTCATTGGCGTGAATGACCTTGAAAGGGAGGGACAGTACATGTTCACAGACAACACTCCA 5994

Qy    687 ctgcagaactatagcaactggaatgaggggggaacccagcgacccctatggtcatgaggac 746
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5993 CTGCAGAACTATAGCAACTGGAATGAGGGGGAACCCAGCGACCCCTATGGTCATGAGGAC 5934

Qy    747 tgtgtggagatgctgagctctggcagatggaatgacacagagtgccatcttaccatgtac 806
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5933 TGTGTGGAGATGCTGAGCTCTGGCAGATGGAATGACACAGAGTGCCATCTTACCATGTAC 5874

Qy    807 tttgtctgtgagttcatcaagaagaaaaagtaacttccctcatcctacgtatttgctatt 866
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5873 TTTGTCTGTGAGTTCATCAAGAAGAAAAAGTAACTTCCCTCATCCTACGTATTTGCTATT 5814

Qy    867 ttcctgtgaccgtcattacagttattgttatccatcctttttttcctgattgtactacat 926
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5813 TTCCTGTGACCGTCATTACAGTTATTGTTATCCATCCTTTTTTTCCTGATTGTACTACAT 5754

Qy    927 ttgatctgagtcaacatagctagaaaatgctaaactgaggtatggagcctccatcatcat 986
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5753 TTGATCTGAGTCAACATAGCTAGAAAATGCTAAACTGAGGTATGGAGCCTCCATCATCAT 5694

Qy    987 gctcttttgtgatgattttcatattttcacacatggtatgttattgacccaataactcgc 1046
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5693 GCTCTTTTGTGATGATTTTCATATTTTCACACATGGTATATTATTGACCCAATAACTCGC 5634

Qy   1047 caggttacatgggtcttgagagagaattttaattactaattgtgcacgagatagttggtt 1106
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5633 CAGGTTACATGGGTCTTGAGAGAGAATTTTAATTACTAATTGTGCACGAGATAGTTGGTT 5574

Qy   1107 gtctatatgtcaaatgagttgttctcttggtatttgctctaccatctctccctagagcac 1166
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5573 GTCTATATGTCAAATGAGTTGTTCTCTTGGTATTTGCTCTACCATCTCTCCCTAGAGCAC 5514

Qy   1167 tctgtgtctatcccagtggataatttcccagtttactggtgatgattaggaaggttgttg 1226
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5513 TCTGTGTCTATCCCAGTGGATAATTTCCCAGTTTACTGGTGATGATTAGGAAGGTTGTTG 5454

Qy   1227 atggttaggctaacctgccctggcccaaagccagacatgtacaagggctttctgtgagca 1286
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5453 ATGGTTAGGCTAACCTGCCCTGGCCCAAAGCCAGACATGTACAAGGGCTTTCTGTGAGCA 5394

Qy   1287 atgataagatctttgaatccaagatgcccagatgttttaccagtcacaccctatggccat 1346
          ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db   5393 ATGATAAGATCTTTGAATCCAAGATGCCCAGATCTTTTACCAGTCACACCCTATGGCCAT 5334

Qy   1347 ggctatacttggaagttctccttgttggcacagacatagaaatgctttaaccccaagcct 1406
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db   5333 GGCTATACTTGGAAGTTCTCCTTGTTGGCACAGACATAGAAATGCTTTAACCCCAAACCT 5274

Qy   1407 ttatatgggggacttctagctttgtgtcttgtttcagaccatgtggaatgataaatactc 1466
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5273 TTATATGGGGGACTTCTAGCTTTGTGTCTTGTTTCAGACCATGTGGAATGATAAATACTC 5214

Qy   1467 tttttgtgcttctgatctatcgatttcactaacatataccaagtaggtgctttgaacccc 1526
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5213 TTTTTGTGCTTCTGATCTATCGATTTCACTAACATATACCAAGTAGGTGCTTTGAACCCC 5154

Qy   1527 tttctgtaggctcacaccttaatctcaggcccctatatagtcacactttgatttaagaaa 1586
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5153 TTTCTGTAGGCTCACACCTTAATCTCAGGCCCCTATATAGTCACACTTTGATTTAAGAAA 5094

Qy   1587 aacggagc 1594
          || |||||
Db   5093 AATGGAGC 5086
```

```
RESULT 14
AF282902
LOCUS       AF282902               5851 bp    mRNA    linear   INV 17-DEC-2000
DEFINITION  Hydra vulgaris type IV collagen alpha 1 chain precursor, mRNA,
            complete cds.
ACCESSION   AF282902
VERSION     AF282902.1  GI:11875611
KEYWORDS    .
SOURCE      Hydra vulgaris.
  ORGANISM  Hydra vulgaris
            Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Anthomedusae;
            Hydridae; Hydra.
REFERENCE   1  (bases 1 to 5851)
  AUTHORS   Fowler,S.J., Jose,S., Zhang,X., Deutzmann,R., Sarras,M.P. Jr. and
            Boot-Handford,R.P.
  TITLE     Characterization of hydra type IV collagen. TYPE IV COLLAGEN IS
            ESSENTIAL FOR HEAD REGENERATION AND ITS EXPRESSION IS UP-REGULATED
            UPON EXPOSURE TO GLUCOSE
  JOURNAL   J. Biol. Chem. 275 (50), 39589-39599 (2000)
  MEDLINE   20564332
REFERENCE   2  (bases 1 to 5851)
  AUTHORS   Fowler,S.J. and Boot-Handford,R.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JUN-2000) School of Biological Sciences, University
            of Manchester, Oxford Road, Manchester M13 9PT, UK
FEATURES             Location/Qualifiers
     source          1. .5851
                     /organism="Hydra vulgaris"
                     /db_xref="taxon:6087"
     sig_peptide     157. .228
     CDS             157. .5328
                     /codon_start=1
                     /product="type IV collagen alpha 1 chain precursor"
                     /protein_id="AAG40729.1"
                     /db_xref="GI:11875612"
                     /translation="MNHTTKWTIWIVLIFTLHSCLVFTQLCGGGGCNQCSNVCVGQKG
                     DRGSVGLPGFKGPIGEPGFPGGEGPAGRPGEKGDGGAPGLQGEVGARGKVGPPGTPGI
                     PGSHGRPGDEGRQGEPGLPGCNGTKGDIGPPGPRGKDGVQGPSGLPGPIGPPGEPGDS
                     TATKLKGQKGEPGPKSQDGPAGDPGTKGDKGESGKAGRMGPQGLRGEKGARGDSNITI
                     FGERGDKGDIGLPGPPGRDCNGSSVTGELITNIQGPKGEQGKKGDQGQKGEPGQPGQA
                     GEAGQDGQKGEKGDKGEIGSAGPSGIQGEKGDIGATGAAGTKGEPGSDGSLGSPGQKG
                     DHGTKGSEGPVGDRGKQGESITGPPGQIGEKGQRGPEGKKGEPGPEGPPGPNGEVGDT
                     GPPGLKGLKGEIGMTGPSGEPGKTGAEGMKGPIGPAGQRGETGSKGESGRPGQSVQGS
                     PGMDGNPGQIGPPGEPGIPGFPGPPGNPGQVVNDLGETVLPGPPGDQGPQGIQGVAGP
                     QGSPGIRGDKGDACKSCPSGPQGEKGVAGQDGLPGANGEKGDKGDEGKPGLEGSVGDP
                     GESGLTGPQGPIGEKGDTGAKGNKGERGSDRIVQGERGEKGTDGMPGSPGSIGFKGEQ
                     GVPGDPGTDGPQGPQGLQGPQGDRGAPGKSGIPGNNGEKGEKGNKGDGLRGFKGERGR
                     DGEPGPKGSQGPMGKPGPKGEPFSTALKGQQGDKGDVGSEGPPGPQGAEGQVGNQGFK
                     GEPGLRGAPGSGEKGEKGDAGTPGIAGKNGENGKDGQDGPKGDKGEKGLQGPIGISGP
                     IKGEPGLKGDPGKPGLDGAPGKSVPKGEPGRDGKDGDPGAKGEPGTPGESGRDGEKGE
                     IGPQGPQGDKGDQGQSVIGPKGETGAIGKKGETGQDGLKGRDGIGKQGANGEKGEKGN
                     SVVGLPGDPGEKGSPGEKGKDGISGKQGKPGPPGKDGVDGTPGKDGLPGNPGPEGPKG
                     EKGVTGPTGLDGEKGEPGYSGQKGAKGETGVRGKPGATSNVMGPKGNRGESGAKGEPG
                     KQGPPGAEGPQGEKGERGIGTDKGTKGDTGDIGLQGPKGERGERGEQGRKGDQGPIGP
                     KGDVGEPGPAGSIGNTGISGNDGPKGESGPPGIKGEPGSVGSIGPKGSKGEQGEKGER
                     GSTLKGNPGPEGPRGPTGPPGPAGKDYILDLEVGPLGEKPTKGEPGEDGPKGDKGDIG
                     ETGSQGTKGEKGVPGSKGEKGEQGLEGPKGLEGRVGLAGSKGDKGIKGVIGKPGPTGD
                     KGEKGLQGFKGDTGPQGIQGKEGEPGREGAKGLKGEVGAEGKMGPPCDAGATIIGEKG
                     DRGEKGEAGQRGYKGEPGPSSQLTINPGAKGEQGDRGIKGEKGNIGPKGINGPFGKDG
                     SIGPEGPKGDSGQKGETGLDGPKGEKGQPAIGTGPKGEKGDQGPPGPQGPQGDIGEKG
                     DIGPPGETGDIGPQGIKGEFGIAGIKGETGPPGLQGPIGLKGDQGPPGDAGKVGPPGR
                     SGSGDFGFYLVKHSQSIKVPSCPAGMQTMWEGYSFLYAQGNERAFGQDLGQPGSCLKR
                     FSTMPFLFCDIQNKCVVASRNDYSFWLSTAEKPKEAPSSGADLENYISRCIVCEAPSH
                     VLAVHSQSELDPKCPDGWENLWTGFSFLMYNSAGAQGSGQLLSSSGSCLEDFRVNPYI
                     ECHGRGTCWYYGPTLSFWLSTIGESNMFQVPKFEILERNLKARVSRCAVCMKSVP"
     misc_feature    229. .276
                     /note="Region: non-collagenous domain"
     misc_feature    277. .4641
                     /note="Region: contains collagenous domain"
     misc_feature    526. .528
                     /note="N-glycosylation; glycosylation site"
                     /evidence=not_experimental
     misc_feature    4642. .5325
                     /note="Region: non-collagenous domain"
BASE COUNT     1953 a    942 c   1638 g   1318 t
ORIGIN
```

```
  Query Match             4.6%;  Score 74;  DB 3;  Length 5851;
  Best Local Similarity   58.0%;  Pred. No. 3.6e-08;
  Matches  131;  Conservative    0;  Mismatches   95;  Indels    0;  Gaps    0;

Qy    132 acaatttcaccaggacccaaaggagatgatggtgaaaaaggagatccaggagaagaggga 191
          |||| |   | ||| || ||||| ||  | ||| |||||||||||| |||| || | |||
Db    901 ACAAATATTCAAGGCCCAAAAGGTGAGCAAGGTAAAAAAGGAGATCAAGGACAAAAAGGA 960

Qy    192 aagcatggcaaagtgggacgcatggggccgaaaggaattaaaggagaactgggtgatatg 251
            | |  ||  ||   || |     ||       ||     | ||  ||   || || |
Db    961 GAACCAGGACAACCTGGTCAAGCTGGAGAAGCGGGTCAGGATGGTCAAAAAGGAGAAAAA 1020

Qy    252 ggagatcggggcaatattggcaagactgggcccattgggaagaagggtgacaaaggggaa 311
          || |||   ||  | ||||| |   |||| ||   ||| |   | ||||| ||||| ||
Db   1021 GGTGATAAAGGAGAGATTGGTAGTGCTGGACCTTCTGGAATTCAAGGTGAAAAAGGTGAT 1080

Qy    312 aaaggtttgcttggaatacctggagaaaaaggcaaagcaggtactg 357
          |  |||     |||   | ||||   ||||||  || ||||  |||
Db   1081 ATTGGTGCAACTGGTGCAGCTGGTACAAAAGGGGAACCAGGGTCTG 1126
```

```
RESULT 15
SPY301812
LOCUS       SPY301812              1529 bp    DNA     linear   BCT 11-MAR-2001
DEFINITION  Streptococcus pyogenes sclB gene, regulatory region, strain 684.
ACCESSION   AJ301812
VERSION     AJ301812.1  GI:13235594
KEYWORDS    sclB gene.
SOURCE      Streptococcus pyogenes.
  ORGANISM  Streptococcus pyogenes
            Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
            Streptococcus.
REFERENCE   1  (bases 1 to 1529)
  AUTHORS   Whatmore,A.M.
  TITLE     Streptococcus pyogenes sclB encodes a putative hypervariable
            surface protein with a collagen-like repetitive structure
  JOURNAL   Microbiology 147 (Pt 2), 419-429 (2001)
   PUBMED   11158359
REFERENCE   2  (bases 1 to 1529)
  AUTHORS   Whatmore,A.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-NOV-2000) Whatmore A.M., Biological Sciences,
            University Of Warwick, Coventry, CV4 7AL, UNITED KINGDOM
FEATURES             Location/Qualifiers
     source          1. .1529
                     /organism="Streptococcus pyogenes"
                     /strain="684"
                     /db_xref="taxon:1314"
     -35_signal      36. .41
                     /gene="sclB"
     gene            36. .1497
                     /gene="sclB"
     -10_signal      60. .65
                     /gene="sclB"
     repeat_region   146. .195
                     /rpt_family="CAAAA"
                     /rpt_unit=146. .150
     terminator      1474. .1497
                     /gene="sclB"
BASE COUNT      541 a    324 c    379 g    285 t
ORIGIN
```

```
  Query Match             4.6%;  Score 73.6;  DB 1;  Length 1529;
  Best Local Similarity   49.8%;  Pred. No. 4.6e-08;
  Matches  214;  Conservative    0;  Mismatches  214;  Indels    2;  Gaps    1;

Qy     89 ggatattgatagccgtcctaccgctgaagtctgtgccacacacacaatttcaccaggacc 148
```

```
Db    619 GGGTGCAGATGGAGCGCCAGGTGCCAAAGGTGATCGCGGCGAAGCCGGTCCTGCAGGCCC 678

Qy    149 caaaggagatgatggtgaaaaaggagatccaggagaagagggaaagcatggcaaagtggg 208

Db    679 AGTAGGTCCACGTGGTGACAAGGGTGAAAAAGGAGAACAAGGAGAACAAGGCCTAAGGGG 738

Qy    209 acgcatggggccgaaaggaattaaaggagaactgggtgatatgggagatcggggcaatat 268

Db    739 TGAACCTGGAGAAAAGGGTGCAGATGGAGCGCCAGGTGCCAAAGGTGATCGCGGTGAAAC 798

Qy    269 tggcaagactgggcccattgggaagaagggtgacaaaggggaaaaaggtttgcttggaat. 328

Db    799 CGGCCCAGTAGGCCCACGTGGTGACAAGGGTGAAAAAGGAGCTCAAGGTCCCGCCGGAAA 858

Qy    329 acctggagaaaaaggcaaagcaggtactgtctgtgattgtggaagataccggaaatttgt 388

Db    859 AGATGGTGAACGTGGTCCAGTAGGTCCCGCTGGCAAGGACGGCCAAGATGGCAAAGATGG 918

Qy    389 tggacaactggatattagtattgcccggctcaagacatctatgaagtttgtcaagaatgt 448

Db    919 TCTTCCAGGTAAAGACGGTAAGGAC--GGCCAAGATGGCAAAGATGGTCTTCCAGGTAAA 976

Qy    449 gatagcagggattagggaaactgaagagaaattctactacatcgtgcaggaagagaagaa 508

Db    977 GACGGCAAGGACGGCCAAAACGGTAAAGATGGTCTCCCAGGTAAAGACGGTAAGGACGGC 1036

Qy    509 ctacagggaa 518

Db   1037 CAAGATGGCA 1046
```

Search completed: July 6, 2002, 16:03:34
Job time: 49655 sec